

Lecture Notes in Artificial Intelligence 2685

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Springer

Berlin

Heidelberg

New York

Barcelona

Hong Kong

London

Milan

Paris

Tokyo

Christian Freksa Wilfried Brauer
Christopher Habel Karl F. Wender (Eds.)

Spatial Cognition III

Routes and Navigation,
Human Memory and Learning,
Spatial Representation and Spatial Learning



Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Christian Freksa
Universität Bremen, FB 3 - Mathematik und Informatik
Bibliothekstr. 1, 28359 Bremen, Germany
E-mail: freksa@sfbtr8.uni-bremen.de

Wilfried Brauer
Technische Universität München, Fakultät für Informatik
Boltzmannstr. 3, 85748 Garching bei München, Germany
E-mail: brauer@informatik.tu-muenchen.de

Christopher Habel
Universität Hamburg, Fachbereich Informatik
Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
E-mail: habel@informatik.uni-hamburg.de

Karl F. Wender
Universität Trier, FB 1 - Psychologie
54286 Trier, Germany
E-mail: wender@cogpsy.uni-trier.de

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek.
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

CR Subject Classification (1998): I.2.4, I.2, J.2, J.4, E.1, I.3, I.7, I.6

ISSN 0302-9743

ISBN 3-540-40430-9 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2003
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Steingraber Satztechnik GmbH, Heidelberg
Printed on acid-free paper SPIN: 10927816 06/3142 5 4 3 2 1 0

Preface

Spatial cognition is an interdisciplinary research area involving artificial intelligence, cognitive psychology, computational linguistics, geography, mathematics, biology, theoretical computer science, architecture, design, and philosophy of mind. As these different disciplines gain a deeper understanding of their fellow disciplines and their research approaches, they increasingly find ways to combine their insights and to conceive powerful mechanisms to analyze and synthesize cognitive systems. Spatial cognition has now reached a point where we can see how different pieces of the puzzle may fit together to form integrated systems of specialized cognitive components. The research triggers new quests for basic issues of cognition and sparks ideas for the development of technological applications that make use of spatial structures and spatial computation. Potential applications can be found in such diverse areas as autonomous robotics, geographic information systems, location-based services, spatial task assistance, multi-agent communication, to name but a few.

This third volume on *Spatial Cognition* marks the final phase of the German Spatial Cognition Priority Program. It augments the results presented in the two precursor volumes published in 1998 and 2000, respectively. The interdisciplinary research program¹ was established by the Deutsche Forschungsgemeinschaft (DFG) in 1996 and terminated after six years, the maximum duration of DFG priority programs. The Spatial Cognition Priority Program consists of 17 research projects at 13 research institutions throughout Germany. Besides carrying out research in individual projects and joint research between projects, the program organized ‘topical colloquia’ and annual plenary colloquia, largely with international participation.

The present volume consists of revised contributions to the eighth plenary colloquium of the Spatial Cognition Priority Program, *Spatial Cognition 2002*, which was held at the Evangelische Akademie in Tutzing (Bavaria) 20–23 May 2002. Topics addressed include diagrammatic representation; spatial ontology, geometry, and paronomies; cognitive robotics; spatial reference systems; spatial reasoning; navigation; geoinformation; spatial memory; knowledge acquisition, imagery, and motion; and virtual reality. The contributions were peer-reviewed before the conference and carefully revised afterwards.

We would like to thank all participants of *Spatial Cognition 2002* and all authors for their contributions and for their revisions in accordance with the reviewers’ recommendations. We thank our commentators and reviewers for their insightful and thorough reviews. We are indebted to Thora Tenbrink for her superb editorial support. We thank the LNAI Series editors Jaime G. Carbonell and Jörg Siekmann as well as Alfred Hofmann of Springer-Verlag for supporting this publication project. We gratefully acknowledge the support of the Spatial Cognition Priority Program by the Deutsche Forschungsgemeinschaft. We thank the members of the review

¹ See www.spatial-cognition.de

committee, Herbert Heuer, Elke van der Meer, Manfred Pinkal (chair), Michael M. Richter, Dirk Vorberg, Ipke Wachsmuth, and Wolfgang Wahlster for their guidance and their support. We are indebted to Andreas Engelke and Gerit Sonntag for their dedicated administration of our research program and for their valuable advice. We acknowledge the support by Erna Büchner and Katja Fleischer of the DFG. We thank Hildegard Westermann of the Knowledge and Language Processing Group at the University of Hamburg for her continuous support of the Spatial Cognition Priority Program. Finally, we wish to thank the Evangelische Akademie Tutzing for providing a stimulating and productive conference environment and for the hospitality they provided for the five plenary meetings we have held at their conference center. In particular, we are indebted to Renate Albrecht of the Akademie Tutzing for accommodating all our special requests and making us feel at home in Schloss Tutzing.

March 2003

Christian Freksa
Wilfried Brauer
Christopher Habel
Karl F. Wender

Commentators and Reviewers

Gary Allen
Elisabeth André
Philippe Balbiani
Jürgen Bohnemeyer
Anthony Cohn
Carola Eschenbach
Klaus Eyferth
Petra Jansen
Karl Gegenfurtner
Daniel Hernández
Stephen Hirtle
Bernhard Hommel
Robin Hörnig
Markus Knauff
Alois Knoll
Werner Kuhn
Lars Kulik
Bernd Leplow
Gérard Ligozat
Gerd Lüer
Hanspeter Mallot
Mark May
Timothy McNamara
Silvia Mecklenbräuer
Daniel Montello

Reinhard Moratz
Lynn Nadel
Bernhard Nebel
Patrick Péruch
Michael Popp
Jochen Renz
Gert Rickheit
Thomas Röfer
Florian Röhrbein
Hedda Schmidtke
Karin Schweizer
Jeanne Sholl
Sibylle Steck
Klaus Stein
John Stell
Thora Tenbrink
Barbara Tversky
Ipke Wachsmuth
Monika Wagener-Wender
Wolfgang Wahlster
Mike Worboys
Steffen Werner
Jianwei Zhang
Hubert Zimmer

Related Book Publications

- Barkowsky, T., *Mental Representation and Processing of Geographic Knowledge. A Computational Approach*, LNAI 2541, Springer, Berlin 2002.
- Egenhofer, M.J.; Mark, D.M., eds., *Geographic Information Science*, LNCS 2478, Springer, Berlin 2002.
- Hegarty, M.; Meyer, B.; Narayanan, N.H., eds., *Diagrammatic Representation and Inference*, LNCS 2317, Springer, Berlin 2002.
- Coventry, K.; Olivier, P., eds., *Spatial language: Cognitive and computational perspectives*, Kluwer, Dordrecht 2002.
- Renz, J., *Qualitative Spatial Reasoning with Topological Information*, LNAI 2293, Springer, Berlin 2002.
- Montello, D.R., ed., *Spatial Information Theory: Foundations of Geographic Information Science*, LNCS 2205, Springer, Berlin 2001.
- Freksa, C.; Brauer, W.; Habel, C.; Wender, K. F., eds., *Spatial Cognition II - Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, LNAI 1849, Springer, Berlin 2000.
- Habel, C.; von Stutterheim, C.; Hrsg., *Räumliche Konzepte und sprachliche Strukturen*. Niemeyer, Tübingen 2000.
- Habel, C.; Werner, S., eds. Special Issue on Spatial Reference Systems. *Spatial Cognition and Computation. Vol 1, No.4* (1999).
- Freksa, C.; Mark, D.M., eds., *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*. LNCS 1661, Springer, Berlin 1999.
- Freksa, C.; Habel, C.; Wender, K.F., eds., *Spatial Cognition*. LNAI 1404, Springer, Berlin 1998.
- Egenhofer, M.J.; Golledge, R.G., eds., *Spatial and Temporal Reasoning in Geographic Information Systems*. Oxford University Press, Oxford 1997.
- Hirtle, S.C.; Frank, A.U., eds. *Spatial Information Theory: A Theoretical Basis for GIS*, LNCS 1329, Springer, Berlin 1997.
- Burrough, P.; Frank, A., eds., *Geographic objects with indeterminate boundaries*, Taylor and Francis, London 1996.
- Frank, A.U.; Kuhn, W., eds. *Spatial Information Theory: A Theoretical Basis for GIS*, LNCS 988, Springer, Berlin 1995.
- Frank, A.U.; Campari, I., eds. *Spatial Information Theory: A Theoretical Basis for GIS*, LNCS 716, Springer, Berlin 1993.
- Frank, A.U.; Campari, I.; Formentini, U., eds., *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, LNCS 639, Springer, Berlin 1992.
- Mark, D.M.; Frank, A.U., eds., *Cognitive and linguistic aspects of geographic space*, 361-372, Kluwer, Dordrecht 1991.
- Freksa, C.; Habel, C., Hrsg., *Repräsentation und Verarbeitung räumlichen Wissens*, Informatik-Fachberichte 245, Springer, Berlin 1990.

Table of Contents

Routes and Navigation

Navigating by Mind and by Body	1
<i>Barbara Tversky</i>	
Pictorial Representations of Routes:	
Chunking Route Segments during Comprehension	11
<i>Alexander Klippel, Heike Tappe, Christopher Habel</i>	
Self-localization in Large-Scale Environments for the Bremen Autonomous Wheelchair	34
<i>Axel Lankenau, Thomas Röfer, Bernd Krieg-Brückner</i>	
The Role of Geographical Slant in Virtual Environment Navigation	62
<i>Sibylle D. Steck, Horst F. Mochnatzki, Hanspeter A. Mallot</i>	
Granularity Transformations in Wayfinding	77
<i>Sabine Timpf, Werner Kuhn</i>	
A Geometric Agent Following Route Instructions	89
<i>Ladina B. Tschander, Hedda R. Schmidtke, Carola Eschenbach, Christopher Habel, Lars Kulik</i>	
Cognition Meets Le Corbusier – Cognitive Principles of Architectural Design	112
<i>Steffen Werner, Paul Long</i>	

Human Memory and Learning

The Effect of Speed Changes on Route Learning in a Desktop Virtual Environment	127
<i>William S. Albert, Ian M. Thornton</i>	
Is It Possible to Learn and Transfer Spatial Information from Virtual to Real Worlds?	143
<i>Doris Höll, Bernd Leplow, Robby Schönfeld, Maximilian Mehdorn</i>	
Acquisition of Cognitive Aspect Maps	157
<i>Bernhard Hommel, Lothar Knuf</i>	
How Are the Locations of Objects in the Environment Represented in Memory?	174
<i>Timothy P. McNamara</i>	

Priming in Spatial Memory: A Flow Model Approach	192
<i>Karin Schweizer</i>	

Context Effects in Memory for Routes	209
<i>Karl F. Wender, Daniel Haun, Björn Rasch, Matthias Blümke</i>	

Spatial Representation

Towards an Architecture for Cognitive Vision	
Using Qualitative Spatio-temporal Representations and Abduction	232
<i>Anthony G. Cohn, Derek R. Magee, Aphrodite Galata,</i> <i>David C. Hogg, Shyamanta M. Hazarika</i>	

How Similarity Shapes Diagrams	249
<i>Merideth Gattis</i>	

Spatial Knowledge Representation for Human-Robot Interaction	263
<i>Reinhard Moratz, Thora Tenbrink, John Bateman, Kerstin Fischer</i>	

How Many Reference Frames?	287
<i>Eric Pederson</i>	

Motion Shapes: Empirical Studies and Neural Modeling	305
<i>Florian Röhrbein, Kerstin Schill, Volker Baier, Klaus Stein,</i> <i>Christoph Zetsche, Wilfried Brauer</i>	

Use of Reference Directions in Spatial Encoding	321
<i>Constanze Vorwerg</i>	

Spatial Reasoning

Reasoning about Cyclic Space: Axiomatic and Computational Aspects	348
<i>Philippe Balbiani, Jean-François Condotta, Gérard Ligozat</i>	

Reasoning and the <i>Visual-Impedance</i> Hypothesis	372
<i>Markus Knauff, P.N. Johnson-Laird</i>	

Qualitative Spatial Reasoning about Relative Position	385
<i>Reinhard Moratz, Bernhard Nebel, Christian Freksa</i>	

Interpretation of Intentional Behavior in Spatial Partonomies	401
<i>Christoph Schlieder, Anke Werner</i>	

Author Index	415

Navigating by Mind and by Body

Barbara Tversky¹

Stanford University
Department of Psychology
420 Jordan Hall
Stanford, Ca 94305
{bt@psych.stanford.edu}

Abstract. Within psychology, at least two research communities study spatial cognition. One community studies systematic errors in spatial memory and judgement, accounting for them as a consequence of and clue to normal perceptual and cognitive processing. The other community studies navigation in real space, isolating the contributions of various sensory cues and sensori-motor systems to successful navigation. The former group emphasizes error, the latter, selective mechanisms, environmental or evolutionary, that produce fine-tuned correct responses.

How can these approaches be reconciled and integrated? First, by showing why errors are impervious to selective pressures. The schematization that leads to errors is a natural consequence of normal perceptual and cognitive processes; it is inherent to the construction of mental spaces and to using them to make judgments in limited capacity working memory. Selection can act on particular instances of errors, yet it is not clear that selection can act on the general mechanisms that produce them. Next, in the wild, there are a variety of correctives. Finally, closer examination of navigation in the wild shows systematic errors, for example, over-shooting in dead reckoning across species. Here, too, environments may provide correctives, specifically, landmarks. General cognitive mechanisms generate general solutions. The errors inevitably produced may be reduced by local specific sensori-motor couplings as well as local environmental cues. Navigation, and other behaviors as well, are a consequence of both.

1 Two Research Communities in Psychology

Yes, the title evokes the mind-body problem. However one regards the venerable monumental mind-body problem in philosophy, there is a contemporary minor mind-body problem in the psychological research on spatial cognition. While the major

¹ I am grateful to Christian Freksa for helpful comments and encouragement and to two anonymous reviewers for critiques of an earlier version of this manuscript. Preparation of the manuscript was supported by Office of Naval Research, Grants Number N00014-PP-1-O649 and N000140110717 to Stanford University.

problem is how to integrate the mind and the body, an additional minor problem in spatial cognition is how to integrate the approaches – and the researchers – on the mind and on the body. The community studying spatial judgments and that studying wayfinding rarely interact. Or have rarely interacted. These conferences of minds may be a meeting point and a turning point.

The two communities, the mind community, and the body community, differ in their agendas and differ in the tools to carry them out. The mind community studies spatial judgments: what is the direction between San Diego and Reno? How far is Manchester from Glasgow? Manchester from Liverpool? The Eiffel Tower to Jacque's house? How do I get to San Marco? The questions are cleverly chosen. They are designed to yield errors. The design works because the errors are a consequence of the way spatial information is represented and used. In fact, one goal of this approach is to reveal those cognitive representations and mechanisms, many of which appear not only in spatial judgments, but in other domains as well (e. g., Tversky, 1993; 2000a; 2000b).

In contrast, the body community studies the cues, visual, auditory, kinesthetic, vestibular, that people and animals use to arrive at their destinations. The research reduces the sensory input and diminishes the environmental richness in order to isolate the role of a particular cue or system in guiding the organism. In many cases, the goal is to reveal the elegant fine-tuning of a particular cue or sets of cues or sensory-motor systems to specific aspects of environments (see, for examples, Gallistel, 1990 and papers in the volume edited by Golledge, 1999, especially the paper by Berthoz, Amorim, Glassauer, Grasso, Takei, and Viaud-Delmon and Loomis, Klatzky, Golledge, and Philbrick).

To caricature the approaches, the emphasis of the mind community is to reveal the systems generating error and the emphasis of the body community is to reveal the systems generating precision.

No wonder the community of mind and the community of body pass each other by like the proverbial ships in the night. They differ in the tasks they give, in the responses they collect, in the processes they propose to account for the responses to the tasks. And, perhaps most significantly, they differ philosophically, in their fundamental attitudes toward human nature. For the mind group, being human is fundamentally about limitations, limitations in representations and in processing, in capacity and in computation. Those limitations can be revealed in errors. The errors provide clues to normal operations. For the body group, being human is fundamentally about evolution and learning, about selection and adaptation, pressures toward perfection. Again, these are caricatures of the positions, hence not attributed to any of the fine reasonable people in the fields, but caricatures that are close enough to the truth to warrant further discussion. And perhaps, rapprochement, even integration, of the approaches.

Neither evolution nor adaptation are doubted. Both communities believe that organisms have evolved in and continue to live in environments, and that the environments have selected successful behaviors across the millennia through evolution and across the lifespan through learning. So the real puzzle is not why some spatial behaviors are exquisitely precise and fine-tuned, but rather why systematic

errors persist. Before that question can be addressed, a review of some of the documented errors is in order. Then these errors must be accounted for by an analysis of the general mechanisms that produce and maintain them.

2 Systematic Distortions of Distance and Direction

2.1 Errors of Distance

First, what errors do we mean? Errors of distance estimates, for one. They are affected by irrelevant factors, such as hierarchical organization. Elements, like cities or buildings, within the same group are perceived as closer than those in different groups. The groups might be states or countries. The groups need not be geographic; they can be functional or conceptual. Distances between a pair of academic buildings or a pair of commercial buildings in Ann Arbor are perceived as shorter relative to distances between an academic and a commercial building (Hirtle and Jonides, 1981). Arabs perceive distances between pairs of Arab settlements to be smaller than distances between an Arab and a Jewish settlement; similarly, Jews perceive distances between Jewish settlements to be shorter than distances between an Arab and a Jewish settlement (Portugali, 1993). Grouping is reflected in reaction times to make distance estimates as well; people are faster to verify distances between geographic entities such as states or countries than within the same entity (e. g., Maki, 1981; Wilton, 1979). Another factor distorting distance estimates is the amount of information along the route. Distance judgments for routes are judged longer when the route has many turns (e. g., Sadalla and Magel, 1980) or landmarks (e. g., Thorndyke, 1981) or intersections (e. g., Sadalla and Staplin, 1980). Similarly, the presence of barriers also increases distance estimates (e. g., Newcombe and Liben, 1982). Most remarkably, distance judgements are not necessarily symmetric. Distances to a landmark are judged shorter than distances from a landmark to an ordinary building (Sadalla, Burroughs, and Staplin, 1980; McNamara and Diwadker, 1997). Similar errors occur for prototypes in similarity judgments: people judge atypical magenta to be more similar to prototypic red than red to magenta (Rosch, 1975). Landmarks seem to define neighborhoods and prototypes categories whereas ordinary buildings and atypical examples do not. Ordinary buildings in the vicinity of a landmark may be included in the neighborhood the landmark defines.

2.2 Errors of Direction

Systematic errors occur for judgments of direction as well. Hierarchical organization is again a factor. For example, the overall direction between pairs of states appears to be used to judge the direction between pairs of cities contained in the states. The example so famous that it has become a Trivial Pursuit question is the direction between San Diego and Reno. Students in San Diego erroneously indicated that San Diego is west of Reno (Stevens and Coupe, 1978). That is, the overall direction of the

states is used to infer the directions between cities within those states. But errors of direction occur within groups as well, for example, informants incorrectly report that Berkeley is east of Stanford (Tversky, 1981). This error seems to be due to mentally rotating the general direction of the surrounding geographic entity, in this case, the south Bay Area to the overall direction of the frame of reference, in this case, north-south. In actuality, the south Bay Area runs nearly diagonally with respect to the overall frame of reference, that is, northwest to southeast. Geographic entities create their own set of axes, typically around an elongated axis or an axis of near symmetry. The axes induced by the region may differ from the axes of its' external reference frame. Other familiar cases include South America, Long Island, Japan, and Italy. In this error of rotation, the natural axes of the region and those of the reference frame are mentally brought into greater correspondence. Directions also get straightened in memory. For example, asked to sketch maps of their city, Parisians drew the Seine as a curve, but straighter than it actually is (Milgram and Jodelet, 1976). Even experienced taxi drivers straighten the routes they ply each day in the maps they sketch (Chase and Chi, 1981).

2.3 Other Errors

These are not the only systematic errors of spatial memory and judgment that have been documented; there are others, notably, errors of quantity, shape, and size, as well as errors due to perspective (e. g., Tversky, 1992; Poulton, 1989). Analogous biases are found in other kinds of judgements: for example, people exaggerate the differences between their own groups, social or political, and other groups, just as they exaggerate the distances between elements in different geographic entities relative to elements in the same geographic entity. The errors are not random or due solely to ignorance; rather they appear to be a consequence of ordinary perceptual and cognitive processes.

3 Why Do Errors Exist?

3.1 Schematization Forms Mental Representations

A number of perceptual and cognitive processes are involved in establishing mental representations of scenes or depictions, such as maps. Isolating figures from grounds is one of them; figures may be buildings or roads, cities or countries, depending on what is represented. Figures are then related to one another and to a frame of reference from a particular perspective (e. g., Tversky, 1981; 1992; 2000a). Natural as they are, essential as they are, these perceptual organizing principles are guaranteed to produce error. They simplify, approximate, omit, and otherwise schematize the geographic information. Schematization thereby produces error.

How does this happen? Consider these examples. Relating figures to one another draws them closer in alignment in memory than they actually are. Evidence comes

from a task where students were asked to select the correct map of the Americas from a pair of maps in which one was correct and the other had been altered so that South America was more aligned with North America. A majority of students selected the more aligned map as the correct one (Tversky, 1981). The same error was obtained for maps of the world, where a majority preferred an incorrect map in which the U.S. and Europe were more aligned. Alignment occurred for estimates of directions between cities, for artificial maps, and for blobs. Relating a figure to a reference frame yields the rotation errors described in the section on errors of direction. Like alignment, rotation occurs for directions between cities, for artificial maps, and for blobs.

3.2 Schematization Allows Integration

Many environments that we know, navigate, and answer questions about are too large to be perceived from a single point. Acquiring them requires integrating different views as the environment is explored. Even perceiving an environment from a single point requires integration of information, from separate eye fixations, for example. How can the different views be integrated? The obvious solution is through common elements and a common reference frame. And these, elements and reference frames, are exactly the schematizing factors used in scene perception. To make matters more complex, knowledge about environments comes not just from exploration, but from maps and descriptions as well, so the integration often occurs across modalities. Again, the way to link different modalities is the same as integrating different views, through common elements and frames of reference.

3.3 Schematization Reduces Working Memory Load

A third reason for schematization is that the judgments are performed in working memory, which is limited in capacity (e. g., Baddeley, 1990). Providing the direction or distance or route between A and B entails retrieving the relevant information from memory. This is unlikely to be in the form of a prestored, coherent memory representation, what has been traditionally regarded as a cognitive map. More likely it entails retrieving scattered information and organizing it. Moreover, whatever is stored in memory has already been schematized. All this, and the judgment as well, is accomplished in working memory. Like mental multiplication, this is burdensome. Anything that reduces load is useful, and schematization does just that. This is similar to reducing bandwidth by compression, but in the case of constructing representations in working memory, the compression is accomplished by schematization, by selecting the features and relations that best capture the information

3.4 Spatial Judgments Are Typically Decontextualized

Unlike navigation by the body, navigation in the mind is without support of context. This is in sharp contrast to the spatial behaviors that are precise, accurate, and finely-tuned, such as catching balls, playing the violin, wending one's way through a crowd, finding the library or the subway station. Context provides support in several ways. First it provides constraints. It excludes many behaviors and encourages others. The structure of a violin constrains where the hands, fingers, chin can be placed and how they can be moved. The structure of the environment constrains where one can turn, where one can enter and exit. The world does not allow many behaviors that the mind does. Second, natural contexts are typically rich in cues to memory and performance. For memory, contexts, like menus on computer screens, turn recall tasks into recognition tasks. A navigator doesn't need to remember exactly where the highway exit or subway entrance is as the environment will mark them. The presence of context means that an overall plan can leave out detail such as exact location, direction, and distance. In fact, route directions and sketch maps leave out that level of detail, yet have led to successful navigation across cultures and across time (e. g., Tversky and Lee, 1998, 1999). For performance, context facilitates the specific actions that need to be taken. In the case of playing the violin, this includes time and motion, the changing positions of the fingers of each hand. In the case of wayfinding, this also includes time and motion of various parts of the body, legs in walking, arms, hands, and feet in driving.

4 Why Do Errors Persist?

4.1 Rarely Repeated

Context and contextual cues provide one reason why spatial behaviors by the body may be highly accurate and spatial behaviors by the mind biased. Contexts constrain behaviors and cue behaviors. Contexts are also the settings for practice. As any violin player or city dweller knows, the precise accurate spatial behaviors become so by extensive practice. The efforts of beginners at either are full of false starts, error, and confusion. Practice, and even more so, practice in a rich context supporting the behavior, is the exception, not the rule, for navigation by the mind, for judgements from memory. Indeed, for the judgments that we are called upon to make numerous times, we do eventually learn to respond correctly. I now know that Rome is north of Philadelphia and that Berkeley is west of Stanford.

4.2 Learning Is Specific, Not General

But knowing the correct answer to a particular case corrects only that case, it does not correct the general perceptual and cognitive mechanisms that produce schematizations that produce the errors. Knowing that Rome is north of Philadelphia

doesn't tell me whether Rome is north of New York City or Boston. Knowing that Rome is north of Philadelphia doesn't inform me about the direction from Boston to Rio either. Learning is local and specific, not general and abstract. Immediately after hearing an entire lecture on systematic errors in spatial judgments, a classroom of students made exactly the same errors.

The mechanisms that produce the errors are multi-purpose mechanisms, useful for a wide range of behaviors. As noted, the mechanisms that produce errors derive from the mechanisms used to perceive and comprehend scenes, the world around us. The schematizations they produce seem essential to integrating information and to manipulating information in working memory. In other words, the mechanisms that produce error are effective and functional in a multitude of ways.

4.3 Correctives in Context

Another reason why errors persist is that they may never be confronted. Unless I am a participant in some abstruse study, I may never be asked the direction between Rome and Philadelphia, from Berkeley to Stanford. Even if I am asked, I may not be informed of my error, so I have no opportunity to correct it. And if I am driving to Berkeley, my misconception causes me no problem; I have to follow the highways. Similarly, if I think a particular intersection is a right-angle turn when in fact it is much sharper, or if I think a road is straighter than it is, the road will correct my errors, so I can maintain my misconception in peace. In addition, these errors are independent of each other and not integrated into a coherent and complete cognitive map, so there is always the possibility that errors will conflict and cancel (e. g., Baird, 1979; Baird, Merrill, and Tannenbaum, 1979). Finally, in real contexts, the extra cues not available to working memory become available, both cues from the environment, like landmarks and signs, and also cues from the body, kinesthetic, timing, and other information that may facilitate accuracy and overcome error. In short, schematic knowledge, flawed as it is, is often adequate for successful navigation.

5 Systematic Errors in the Wild

Now the caricature of the communities that has been presented needs refinement. Despite millennia of selection by evolution and days of selection by learning, navigation in the wild is replete with systematic errors. One studied example is path integration. Path integration means updating one's position and orientation while navigating according to the changes in heading and distances traveled, the information about one's recent movements in space (Golledge, 1999, p. 122). A blindfolded navigator traverses a path, turns, continues for a while, and then heads back to the start point. How accurate is the turn to home? Ants are pretty good, so are bees, hamsters, and even people. But all make systematic errors. Bees and hamsters overshoot (Etienne, Maurer, Georgakapoulus, and Griffin, 1999). People overshoot small distances and small turns and undershoot large ones (Loomis, Klatzky,

Golledge, and Philbeck, 1999), a widespread error of judgment (Poulton, 1989). But the situation that induced the errors isn't completely wild; critical cues in the environment have been removed by blindfolding or some other means. In the wild, environments are replete with cues, notably, landmarks, that may serve to correct errors.

6 Implications

How do people arrive at their destinations? One way would be to have a low-level, finely-detailed sequence of actions. But this would only work for well-learned routes in unchanging environments; it wouldn't work for new routes or vaguely known routes or routes that encounter difficulties, such as detours. For those, having a global plan as well as local actions seem useful. These are global and local in at least three senses. Plans are global in the sense of encompassing a larger environment than actions, which are local. Plans are also global in the sense of being general and schematic, of being incompletely specified, in contrast to actions, which are specific and specified. Plans are global in the sense of being amodal, in contrast to actions, which are precise movements of particular parts of the body in response to specific stimuli. A route map is a global plan for finding a particular destination, much as a musical score is a global plan for playing a particular piece on the violin. Neither specifies the exact motions, actions to be taken.

Several approaches to robot navigation have recommended the incorporation of both global and local levels of knowledge (e. g., Chown, Kaplan, and Kortenkamp, 1995; Kuipers, 1978, 1982; Kuipers and Levitt, 1988). The current analysis suggests that global and local levels differ qualitatively. The global level is an abstract schematic plan, whereas the local level is specific sensori-motor action couplings. Integrating the two is not trivial.

The gap between the mind navigators and the body navigators no longer seems so large. True, the focus of the mind researchers is on judgments and the challenge is to account for error and while the focus of the body researchers is on behavior and the challenge is to account for success. Yet, both find successes as well as systematic errors. And in the wild, the correctives to the errors are similar, local cues from the environment.

Systematic errors persist because the systems that produce them are general: they are useful for other tasks and they are too remote to be affected by realization of local, specific errors. Spatial judgment and navigation are not the only domains in which humans make systematic errors. Other accounts have been made for other examples (e. g., Tversky and Kahneman, 1983). It makes one think twice about debates about the rationality of behavior. How can we understand what it means to be rational if under one analysis, behavior seems replete with intractable error but under another analysis, the mechanisms producing the error seem reasonable and adaptive.

References

- Baddeley, A. D. (1990). *Human memory: Theory and practice*. Boston: Allyn and Bacon.
- Berthoz, A., Amorim, M-A., Glassauer, S., Grasso, R., Takei, Y., and Viaud-Delmon, I. (1999). Dissociation between distance and direction during locomotor navigation. In R. G. Golledge (Editor), *Wayfinding behavior: Cognitive mapping and other spatial processes*. Pp. 328-348. Baltimore: Hopkins.
- Bryant, D. J. and Tversky, B. (1999). Mental representations of spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 137-156.
- Baird, J. (1979). Studies of the cognitive representation of spatial relations: I. Overview. *Journal of Experimental Psychology: General*, 108, 90-91.
- Baird, J., Merrill, A., & Tannenbaum, J. (1979). Studies of the cognitive representations of spatial relations: II. A familiar environment. *Journal of Experimental Psychology: General*, 108, 92-98.
- Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scene. *Journal of Memory and Language*, 31, 74-98.
- Bryant, D. J., Tversky, B., and Lanca, M. (2001). Retrieving spatial relations from observation and memory. In E. van der Zee and U. Niskanen (Editors), *Conceptual structure and its interfaces with other modules of representation*. Oxford: Oxford University Press.
- Chase, W. G. & Chi, M. T. H. (1981). Cognitive skill: Implications for spatial skill in large-scale environments. In J. H. Harvey (Ed.), *Cognition, social behavior, and the environment*. Pp. 111-136. Hillsdale, N. J.: Erlbaum.
- Etienne, A. S., Maurer, R., Georgakopoulos, J., and Griffin, A. (1999). Dead reckoning (path integration), landmarks, and representation of space in a comparative perspective. In R. G. Golledge (Editor), *Wayfinding behavior: Cognitive mapping and other spatial processes*. Baltimore: Pp. 197-228. Johns Hopkins Press.
- Franklin, N. and Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 63-76.
- Gallistel, C. R. (1989). Animal cognition: The representation of space, time and number. *Annual Review of Psychology*, 40, 155-189.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge: MIT Press.
- Golledge, R. G. (Editor). (1999). *Wayfinding behavior: Cognitive mapping and other spatial processes*. Baltimore: Johns Hopkins Press.
- Hirtle, S. C. and Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory and Cognition*, 13, 208-217.
- Holyoak, K. J. and Mah, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, 14, 328-352.
- Loomis, J.M., Klatzky, R. L., Golledge, R. G., and Philbeck, J. W. (1999) In R. G. Golledge (Editor). *Wayfinding behavior: Cognitive mapping and other spatial properties*. Pp. 125-151. Baltimore: Johns Hopkins Press.
- Maki, R. H. (1981). Categorization and distance effects with spatial linear orders. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 15-32.
- McNamara, T. P. and Diwadkar, V. A. (1997). Symmetry and asymmetry of human spatial memory. *Cognitive Psychology*, 34, 160-190.
- Milgram, S. and Jodelet, D. (1976). Psychological maps of Paris. In H. Proshansky, W. Ittelson, and L. Rivlin (Eds.), *Environmental Psychology (second edition)*. Pp. 104-124. N. Y.: Holt, Rinehart and Winston.

- Newcombe, N. and Liben, L. (1982). Barrier effects in the cognitive maps of children and adults. *Journal of Experimental Child Psychology*, 34, 46-58.
- Portugali, Y. (1993). *Implicate relations: Society and space in the Israeli-Palestinian conflict*. The Netherlands: Kluwer.
- Poulton, E. C. (1989). *Bias in quantifying judgements*. Hillsdale, N. J.: Erlbaum Associates.
- Rosch, E. (1975). Cognitive reference point. *Cognitive Psychology*, 7, 532-547.
- Sadalla, E. K., Burroughs, W. J., and Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 516-528.
- Sadalla, E. K. and Magel, S. G. (1980). The perception of traversed distance. *Environment and Behavior*, 12, 65-79.
- Sadalla, E. K. and Staplin, L. J. (1980). The perception of traversed distance: Intersections. *Environment and Behavior*, 12 167-182.
- Thorndyke, P. (1981) Distance estimation from cognitive maps. *Cognitive Psychology*, 13, 526-550.
- Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, 90, 293-315.
- Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13, 407-433.
- Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, 23, 131-138.
- Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank and I. Campari (Editors), *Spatial information theory: A theoretical basis for GIS*. Pp. 14-24. Berlin: Springer-Verlag.
- Tversky, B. (2000a). Levels and structure of cognitive mapping. In R. Kitchin and S. M. Freundschuh (Editors). *Cognitive mapping: Past, present and future*. Pp. 24-43. London: Routledge.
- Tversky, B. (2000b). Remembering spaces. In E. Tulving and F. I. M. Craik (Editors), *Handbook of Memory*. Pp. 363-378. New York: Oxford University Press.
- Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Editor), *Spatial schemas and abstract thought*. Pp. 79-111. Cambridge: MIT Press.
- Tversky, B., Kim, J. and Cohen, A. (1999). Mental models of spatial relations and transformations from language. In C. Habel and G. Rickheit (Editors), *Mental models in discourse processing and reasoning*. Pp. 239-258. Amsterdam: North-Holland.
- Tversky, B., & Lee, P. U. (1998). How space structures language. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representation and processing of spatial knowledge* (pp. 157-175). Berlin: Springer-Verlag.
- Tversky, B., & Lee, P. U. (1999). Pictorial and verbal tools for conveying routes. In C., Freksa, & D. M., Mark, (Eds.), *Spatial information theory: Cognitive and computational foundations of geographic information science* (pp. 51-64). Berlin: Springer.
- Wilton, R. N. (1979). Knowledge of spatial relations: The specification of information used in making inferences. *Quarterly Journal of Experimental Psychology*, 31, 133-146.

Pictorial Representations of Routes: Chunking Route Segments during Comprehension

Alexander Klippel, Heike Tappe, and Christopher Habel

University of Hamburg, Department for
Informatics and Cognitive Science Program
[klippel,tappe,habel]@informatik.uni-hamburg.de

Abstract. Route directions are usually conveyed either by graphical means, i.e. by illustrating the route in a map or drawing a sketch-maps or, linguistically by giving spoken or written route instructions, or by combining both kinds of external representations. In most cases route directions are given *in advance*, i.e. prior to the actual traveling. But they may also be communicated quasi-simultaneously to the movement along the route, for example, in the case of in-car navigation systems. We dub this latter kind *accompanying route directions*. Accompanying route direction may be communicated in a dialogue, i.e. with hearer feedback, or, in a monologue, i.e. without hearer feedback. In this article we focus on accompanying route directions without hearer feedback. We start with theoretical considerations from spatial cognition research about the interaction between internal and external representations interconnecting linguistic aspects of verbal route directions with findings from cognitive psychology on route knowledge. In particular we are interested in whether speakers merge elementary route segments into higher order chunks in accompanying route directions. This process, which we identify as *spatial chunking*, is subsequently investigated in a case study. We have speakers produce accompanying route directions without hearer feedback on the basis of a route that is presented in a spatially veridical map. We vary *presentation mode* of the route: In the *static mode* the route is presented as a discrete line, in the *dynamic mode*, it is presented as a moving dot. Similarities across presentation modes suggest overall organization principles for route directions, which are both independent of the type of route direction—in advance versus accompanying—and of presentation mode—static versus dynamic. We conclude that spatial chunking is a robust and efficient conceptual process that is partly independent of preplanning.

Keywords. route map, map-user-interaction, animation, route directions.

1 Internal and External Spatial Representations

The representation of space and the processes that lead to the acquisition of spatial knowledge and its purposeful employment have bothered researchers from various fields of research for the past decades. From an application-oriented point of view, the still growing need to represent and to process spatial knowledge unambiguously arises in areas as diverse as natural language processing, image analysis, visual modeling, robot navigation, and geographical information science. On a theoretical stance,

research has examined the ability of individuals to acquire, use and communicate spatial information as one of our prime cognitive abilities that comprises a wide variety of behavioral competencies and uses a large number of sensory cues, such as kinesthetic, auditory, proprioceptive and visual. Moreover, spatial knowledge may be acquired not only by direct experiential access to an environment but also indirectly: Either by inspecting depictions like photographs, maps, sketches, and virtual computer models or, by exploiting written or spoken descriptions.

In this article we interconnect findings on route knowledge with linguistic findings on verbal route directions. In particular, we focus on a specific process of conceptual organization, namely *spatial chunking*,¹ that combines elementary route segments into higher-order spatial segments (cf. section 2). The hierarchical organization of chunks (Anderson, 1993) is fundamental for hierarchical coding of spatial knowledge (Newcombe & Huttenlocher, 2000). Various kinds of hierarchical structures in the conceptualization of our environment have been investigated in spatial cognition research during the last decades. A starting point of this research is the seminal work of Steven and Coupe (1978). They explore the influence of hierarchical organization on the judgment of spatial relations, namely that a statement like *California is west of Nevada* may lead to misjudgments about the east-west relation with respect to San Diego and Reno. On the other hand, numerous experimental studies provide evidence that and how hierarchical components of spatial memory are basic for efficient and successful spatial problem solving (see, e.g., McNamara, Hardy & Hirtle, 1992). Furthermore, another important aspect of the hierarchical organization of spatial memories is the existence of representations of different degrees or levels of spatial resolution, which can be focused on by mental zooming in and zooming out of representations (cf. Kosslyn 1980).

We investigate the conceptual process of spatial chunking via the analysis of verbal data. Instead of identifying elementary route segments to form a complex sequence of route directions (e.g. *you pass a street to your left but continue walking straight on, then you come to a three-way junction, where again you keep straight on until you come to a branching-off street to your right. Here you turn off.*), they can be combined into a higher order segment (e.g. *you turn to the right at the third intersection*). Thus, a zooming in process makes spatial elements at the lower levels accessible and may result in selecting all decision points for verbalization, whereas zooming out results in spatial chunking and yields higher order segments.

In particular we seek to find out whether spatial chunking is operational during the on-line comprehension of a veridical map² and the verbalization of a route instruction from this map. To this aim we carried out a case study in which participants had to produce a specific sub-type of route direction, namely *accompanying route directions*, which are produced on-line. The route instructions were *accompanying* in that we encouraged the speakers to image a bike-messenger, whom they accompany by giving verbal descriptions via one-way radio messages, i.e. without responses. More

¹ We use the term *chunking* in the tradition of Cognitive Psychology, i.e., referring to a process that builds up *chunks*. We do not make specific theoretic assumptions about the nature of these processes; especially, our usage of *chunking* is not committed to the SOAR approach (Newell, 1990).

² The term veridical map, which contrasts especially to *sketch map*, refers to a map in which focused spatial information is maintained to a high degree. In our case information about distances and angles is preserved.

precisely, the participants were sitting in front of a computer screen displaying a map. They were told to give accurate verbal instructions to a human cyclist traveling through the respective town and thereby directing his movements. They were encouraged to convey the information in such a way that the bike-messenger could follow their instructions without having to ask for clarification. The on-line aspect was enhanced by a *dynamic presentation mode*. In this condition, the route was presented as dot moving through the map leaving the verbalizers little if any cues on the route's continuation. Moreover, we largely impeded preparatory planning processes for both presentation modes: The speakers neither received prior training nor were they presented examples before the actual task. Since we focus on the conceptual chunking processes on part of the *route instructor*³ (rather than the addressee, i.e., the bike-messenger), the accompanying route instructions were given without hearer feedback (cf. section 3 for a detailed description of the setting). If spatial chunking is a general feature in spatial cognition and thus in route directions, the question arises how the presentation mode may affect this conceptual process (cf. Hegarty, 1992; Morrison, Tversky & Betrancourt, 2000).

Route knowledge and verbal route directions have widely been studied from a variety of viewpoints because they provide a richness of empirical cues about the processing of spatial information from different knowledge sources (e.g. Schumacher, Wender & Rothkegel, 2000; Buhl, Katz, Schweizer & Herrmann, 2000; Herrmann, Schweizer, Janzen & Katz, 1998). Route directions are especially apt for investigating the relation between two types of external representations, *graphical* and *linguistic*, and potential intermediary internal representations and principles (cf., e.g. Tversky & Lee, 1999). This is the case as they are usually conveyed either by graphical means—i.e. by illustrating the route in a map or by drawing a sketch-map—or, linguistically—by giving spoken or written route instructions—or by combining both kinds of external representations.

In most cases route directions are given *in advance*, i.e. prior to the addressee's actual action of wayfinding or navigating. In-advance route instructions may be conveyed in situations, which permit different amounts of pre-planning, for example, from 'writing a route instruction for colleagues to help them find the site of a meeting' to 'having to answer the sudden request of a passer-by in a wayfinding situation'. These settings vary according to certain parameters. They have in common, though, that the instructors will start from their spatial knowledge, actually, from that part which regards the requested route. But there are different cognitive tasks to be performed, depending on whether the route instruction is entirely generated from memory, or, in interaction with a map-like representation. In general, spatial cognition research has so far been primarily based on the investigation of spatial representations that are built up from direct experience with the physical world. In most cases the participants were familiar with the environment in question and the empirical investigations were targeted at the participants' long-term memory representations of the respective surrounding, i.e. *spatial mental models* as activated long-term memory representations (Johnson-Laird, 1983) or *cognitive collages* (Tversky, 1993).

In comparison, there are fewer results as to what extent internal representations are built up from external representations of space, namely topographic maps, thematic maps, and sketch-maps and how these representations may differ from those based on

³ Here and in the following we call the speaker who produces the route description, the *route instructor*, or, *instructor* for short.

real-world experience (but see, e.g. Thorndyke & Hayes-Roth, 1982). Generally, the primary role of external representations is their use in solving complex problems by decomposing the representations that are employed in processing the task in external and internal portions (cf. Zhang & Norman, 1994; Zhang, 1997). However, recently, there has been a growing field of research exploring the interaction between external and internal representations (cf. Scaife & Rogers, 1996; Bogacz & Trafton, in press). This also holds for the interaction between map-like representations and spatial cognition (cf. e.g., Barkowsky & Freksa, 1997; Berendt, Rauh & Barkowski, 1998; Casakin, Barkowsky, Klippel & Freksa, 2000; Ghaëm et al., 1998; Hunt & Waller, 1999). In the following sections we review the notions of route knowledge and route directions and explicate our theoretical considerations about the construction of route directions from an external pictorial representation. We clarify the types of external and internal representations in order to specify the spatial chunking processes. Subsequently we present and discuss the results of our case study and conclude with an outlook on future research.

2 External and Internal Representations in Route Knowledge and Spatial Chunking

Route knowledge is characterized as the knowledge about the actions to be performed in the environment to successfully traverse paths between distant locations, especially between an *origin* and a *destination*. Starting from knowledge about landmarks, route learners seem to construct distance and orientation relationships between these fundamental spatial entities and thus come to identify connecting routes between them (Thorndyke & Hayes-Roth, 1982; Golledge, Dougherty & Bell, 1995; Golledge, 1999). Route knowledge is generally assessed by two methods. The first, the *distance estimation task*, where participants have to estimate the distance either between two objects or between themselves and an object. The second, *landmark sequencing*, requires the participants to judge, which of two pictures depicting landmarks located on a route, shows the landmark that would be encountered first coming from a predefined direction. Major features of route knowledge are, first, that it is learned for accomplishing a specific task (mostly, getting from the origin to the destination). Second, that it is based on an egocentric perspective (left and right turns are learned with respect to the body's—actual or imagined—orientation and direction of travel). And third, it is perspective-dependent, meaning that it is most useful when employed from the same viewing perspective as it is learned from (Herrmann, Buhl & Schweizer, 1995). The acquisition of this type of spatial knowledge seems to be primarily based on direct experience.

There is a growing body of research, though, showing that route knowledge can also be acquired from external media (cf. Bell, 1995; Schumacher et al., 2000). For the most part static graphical representations—maps and route sketches—are investigated, while dynamic media for route knowledge learning, like in-vehicle, hand-held and roadside information systems, are still less common. But they are gaining prevalence, which is partly due to the fact that enabling a more efficient distribution of trips over time and space can help limit urban traffic congestion. In parallel to their increasing availability, the cognitive aspects, which underlie the use

of digital navigation aids, receive augmented attention. (Advanced River Navigation, e.g. <http://www.elna.de/de/03/01/01/02/>; Tversky & Lee, 1999; Agrawalla, 2000; Wahlster et al., 2001).

2.1 The Construction of Route Knowledge from External Representations

In the past, maps⁴ were often analyzed as semiotic systems (cf. MacEachren 1995) rather than exploring, how map-users conceptualize the information conveyed in the medium. Yet, recent research has acknowledged that maps are a specific—culturally outstanding—class of external representations that can be characterized by the set of tasks for which maps are regularly applied, namely, spatial problem solving. Particularly, there is a close correspondence between classes of spatial—or more precisely, geographical—problems, on the one hand, and types of maps, on the other hand.

Maps are typically multipurpose means of spatial problem solving: A city map is an external representation to help the user in finding a way from an origin A to a destination B, where A and B span up a variety of potential way finding problems. Even more specialized sketch maps like those designed for finding the way to a specific shopping mall or a chosen hotel are not entirely determined on an individual way finding process: While they are fixed with respect to the destination, they usually make this destination accessible from a (limited) number of origins.

In contrast to such multipurpose external representations for *navigation* and *way-finding* stand specifically tailored means of *way directing*, as verbal route directions, hand drawn sketch maps, or visualizations as well as textual route descriptions produced by computational assistance systems, for example, in car navigation systems.⁵ In the following, we discuss such a type of external representation that is intended for assistance in solving one individual problem, namely giving route directions from an actual origin A to a chosen destination B. In other words, for each pair A and B—constituting a set of routes—a specific route map visualizing the selected route is created and presented to the instructor, whose task it is to simultaneously comprehend and verbalize the route.

This entails that the internal spatial representations of the respective route and its environment, we are concerned with in this paper, are *constructed* rather than inspected during the route direction task. On the one hand, they are therefore likely to resemble the kind of internal representations built up in a concrete navigation situation, where a map is used in order to solve a way-finding problem in an unknown environment. On the other hand, they probably differ from these, in that the instructors are not trying to keep a route or part of it in mind in order to direct their own movements. Rather they give the route instruction while visually sensing the route presented to them in an as yet unknown map. Hence they likely to adhere to the spatial features of the stimulus map because the map itself is veridical and exhibits the spatial layout of the route and its spatial surroundings non-discriminately. In both

⁴ In the following, we use the term *map* generically to refer to various kinds of map-like external representations of space. We will indicate those cases, where a more specific interpretation is intended.

⁵ On these different means of route directing, see, for example, Habel 1988, Freksa 1999, Tversky & Lee 1999.

respects the supposed internal representations for this specific situation might differ from spatial mental models and cognitive collages, which are both considered representations in long-term memory.

2.2 Animation in Pictorial Representations

The major—abstract or rather geometric—property of routes is that they are linear, ordered structures (cf. Eschenbach, Habel & Kulik, 1999; Tschander, Schmidtke, Eschenbach, Habel & Kulik, 2002). The two relevant aspects, namely *linearity* and *ordering*, can be taken into account in map-like representations by different ways of route visualization. Common means are: First, a line, which respects linearity (cf. Figure 2) and second, a line augmented by arrows or arrow-heads, which are conventionalized ways to symbolize orientation of a line. Most recently, dynamic presentations, for example, a distinguished entity moving through the map (cf. sect 3), gain increasing importance in accordance to a growing availability of electronic, stationary, hand-held, and in-car navigation aids. In the case of a dynamically presented route, temporal ordering corresponds to spatial ordering of the route.

In the current paper, we use the first (solid line) and the third (moving dot) means for presenting the stimulus route to the route instructors. The logic behind this juxtaposition is that with the moving-dot condition, i.e. the *dynamic presentation mode*, we enhance the on-line aspect in the verbalization setting. The speakers provide an accompanying route instruction simultaneously to watching the dot moving through the map. In the consequence they might be prone to concentrate on the dot's immediate surrounding which in turn might *discourage* spatial chunking as the chunking process implies the summarization of two or more route segments into one super-ordinate route segment (cf. section 2.3).

With the advent of a growing body of new software tools, current research on diagram understanding has begun to investigate the impact of animated, i.e. dynamically presented, pictorial representations on cognitive processes such as comprehension and reasoning (e.g., Hegarty, 1992). The results are as yet heterogeneous because researchers concentrate both on different kinds of pictorial representations (maps, weather charts, graphs, 3D-forms, etc) and different aspects of cognitive processing (imagery, mental rotation, reasoning, etc.). Thus, there is a range of—reserved to optimistic—estimations about the effects of animation in pictorial representations. While some researchers acknowledge that animation aids in the development of mental models and spatial schema skills for three dimensional forms (Barfield, Lim & Rosenberg, 1990; Augustine & Coover, 1991), others found that animation rather hindered learning and understanding (e.g. Jones & Scaife, 1999; Kaiser, Proffitt, Whelan & Hecht, 1992; Rogers & Scaife, 1997). The latter judgment is based on the finding that animation in pictorial representations often leads to an overload in information, which is hardly integrated into a coherent whole. Morrison et al. (2000) hold that the efficiency of animated graphics is rather doubtful, too. They assert that while animation adds “change over time” to a pictorial representation, this seeming advantage enhances comprehension only in special cases, namely when it succeeds to present micro-steps of processes that static graphics do not present. This finding is akin to the results of Kaiser et al. (1992) who found that even though animation impeded cognitive processing in many cases, it did nonetheless facilitate accurate observation where only *one* dimension of animation was employed.

This exemplary synopsis illustrates, that the question whether and in which way animation influences the comprehension and processing of pictorial representations remains to date unresolved. Furthermore, a universal answer seems unlikely. Rather the impact of animation does most probably depend first, on the specific kind of animation and second, the nature of the cognitive task a particular pictorial representation is designed to assist. The current paper adds to this discussion: We investigate whether there are observable differences in spatial chunking subject to the static or dynamic presentation of the stimulus route in a veridical pictorial representation.

2.3 Route Directions and Spatial Chunking

Verbal route directions are the second distinguished class of external representations to instruct people to find a route. A series of careful analyses from linguistics and psycholinguistics, for example the studies conducted by Denis and his coworkers (viz. Denis, 1997; Denis, Pazzaglia, Cornoldi, & Bertolo, 1998; Daniel & Denis, 1998), provide insights into the mental concepts relevant for route directions.⁶ They put forward the assumption that route instructors can structure their route directions by adhering to the ordering of the spatial objects along the route. Thus, route directions seem to be free from the so-called *linearization problem*, a core problem in language production⁷: “The first remarkable feature of route directions is that they offer a type of spatial discourse in which the linearization problem is not crucial. The object to be described—the route—is not a multidimensional entity but one with an intrinsic linear structure. The discourse simply adheres to the sequence of steps to be followed by the person moving along the route.” (Denis et al., 1999: 147). However, by analyzing great varieties of route directions, Denis et al. (1999) also found that the addressees of route instructions considered very detailed route directions, where every potential *decision point* (i.e. *choice point* or *turn point*) and every landmark was mentioned, rather confusing and rated them to be less appropriate than sparser ones.

From this we conclude that the linearization problem occurs albeit in a slightly different way in that the information encountered in a linear order still has to be organized: Information units can be grouped together and thus a hierarchical structure emerges. For verbalization this hierarchical structure may be traversed at different levels whereby a verbalization of elements at the lowest level corresponds to adhering to the sequence of elements as they appear in temporal order. The verbalization on higher levels of the hierarchy, however, leaves certain elements unmentioned (Habel & Tappe, 1999). In this sense the route instructors are confronted with the central conceptualisation task during language production, namely to detect a *natural order* in the to be described structure and to employ it for verbalization. Since the concept of a ‘natural order’ is extremely vague, one target in modern language production research consists in investigating what kind of ordering is preferable to natural speakers (cf. Tappe, 2000: 71). Applying this principle to route instructions we hold that while the route instructors find it necessary to adhere to the general succession of

⁶ Further aspects are discussed, for example, by Habel 1988; Maaß, 1994; Maaß, Baus & Paul, 1995, Tversky & Lee 1999, and Freksa 1999.

⁷ Linearization means “deciding what to say first, what to say next, and so on” (cf. Levelt, 1989, p.138).

information along the route, it seems preferable to *chunk* some information units—*elementary route segments* in our terminology—together, in order to optimize the amount of information. In route instructions given in advance, spatial chunking and the resulting verbalization of chunked route segments help avoid overload with respect of the addressee's retentiveness, as is exemplified with the contrast between

Turn left at the third intersection

and

You arrive at a crossing, go straight, you pass another branching-off street to your left, do not take this turn, walk straight on until there is a street branching off to your left; Here you turn.

In accompanying route instructions—especially if there is no hearer feedback and the addressee's progression along the route is not entirely transparent to the route instructor—verbalization might not evidence spatial chunking. The instructor might indeed choose to be more detailed in her or his description of the spatial layout and opt to adhere “to the sequence of steps to be followed by the person moving along the route.” Thus, to pinpoint the fundamental difference in the verbalization situation of the participants in our study as compared to the studies of, for example, Denis and his co-workers: The verbalizers in our study have perceptual access to veridical information in form of a map. It is not their memory that determines the elements of the route directions but their conceptualization processes. More importantly even, the route directions are not the results of a planning based process, where the speaker imagines a wellknown environment and mentally constructs a route through this environment which is subsequently verbally conveyed to the addressee. Rather, our participants construct the route directions on-line, while they view the respective map (depicting an unknown environment) for the first time.

2.4 Spatial Chunking and Route Instructions from a Map: External and Internal Representations

In the following we discuss spatial chunking in route instructions via analyzing which kind of information surfaces in verbal route instructions. More specifically, we investigate the question: How does ordering information, i.e. the sequence of graphical-spatial ‘objects’ along the route in the external medium, interact with conceptual processes, especially the spatial chunking of elementary route features? Thus, we have to distinguish between various levels of analysis in the following. On the one hand, we adopt a *medium perspective* to talk about the level of the external representation, i.e. the *map level*. On this level, we find graphical-spatial objects: the signs on the depiction (i.e. map icons) and the graphical structure (i.e. the network of lines representing streets) in which they appear. On the other hand, there are internal representations considered from a *functional perspective*: They are built up for the specific purpose of the route direction and are therefore specific as to the current task. Consequently certain aspects of the external representation, which are—with respect to the task in question—more salient or more important than others, have been transformed from the external representation into internal representations, i.e., they are the primary result of conceptualizing. These internal representations are temporary ‘conceptions’ of the perceived situation. They are both less detailed and less stable than long-term memory representations like spatial mental models or cognitive

collages; they are rather *current spatial representations*. Additionally, under a *procedural perspective* mental processes become apparent that are employed in order to generate functionally adequate route directions.

Table 1. Three perspectives on route directions from maps.

External representation: <i>Medium perspective</i>	Internal representation: <i>Functional perspective</i>	Task-specific processing: <i>Procedural perspective</i>
Spatial Objects: ⁸ - depicted intersections - depicted public locations	Elementary route segments - turning at intersection - landmarks	Chunking: Combination of elementary route segments to elementary and higher order route direction elements

The central question is, what determines the internal representation of a route when a route direction is produced from an external medium? To what extent are route directions the result of human-map interaction? To what extent do they have their own characteristics independently of the specific stimulus? Moreover, can we find differences in processing depending on whether static or dynamic information is processed? And, how do these different kinds of information interact with inherent features of route directions? Similar mechanisms have been discussed for route directions in various environments (Lovelace, Hegarty & Montello, 1999). However, the question of whether the same types of conceptualization are at work when route directions are given from an external medium, such as a map, rather than from a real-world or a simulated environment, has not yet received much attention. Furthermore, whether a variation of the route's presentation mode—static versus dynamic route—has an impact on the spatial chunking is widely unclear. As MacEachren points out: "For dynamic maps and graphs, [...], the fact that time has been demonstrated as indispensable attribute is critical. It tells us that change in position or attributes over time should attract particular attention and serve as a perceptual organizer that is much stronger than hue, value, texture, shape, and so on." (MacEachren, 1995: 35). In the consequence, we suspect the ordering information of graphical-spatial objects along the route to be more salient when the route is presented dynamically to route instructors.

Like in real-world and simulated environments, the information content of the route map is much greater than the information content of the route direction generated from it. During conceptualization 'innocent' map objects become functional route direction features, for example, an intersection is used as a point of a directional change, or an icon for a public location, like a subway station, is employed

⁸ Maps (of this kind) represent real world objects. A distinction can be made between the *object-space* and the *sign-space* (Bollmann, 1993). The term object-space refers to a map without cartographic symbols, i.e. the plain spatial relations of objects are of concern, like in a database. Additionally, for every 'real world' object a cartographic symbol has to be chosen, spanning the sign-space. The salience of an object is not only dependent on its characteristics in the real world (where, for example, a *McDonalds*-restaurant is more salient than a parking lot), it is also dependent on the sign chosen for its representation in the map.

as a landmark. In addition, not every route segment is seized in the same way: Some of them are mentioned explicitly while others are chunked together. In this chunking process elementary route segments are combined, which have, from a non-functional point of view, the same information content as the graphical-spatial objects⁹.

Table 2. Perspectives and route elements.

medium perspective	functional perspective	procedural perspective
crossing 1: three branches (go straight)	CROSSING 1: one relevant branch, no directional change	Chunking of CROSSINGS 2-3
crossing 2: two branches (go straight)	CROSSING 2: one relevant branch, no directional change	
crossing 3: three branches (turn)	CROSSING 3: one relevant branch + directional change TURN	
OUTPUT => turn left at <i>the third crossing</i>		

The chunking process (procedural perspective) accesses elementary route segment (functional perspective), these entities are derived from map objects (external medium perspective), which represent ‘real world’ objects. However, there are other factors, that influence conceptualization, selection of information and linearization (cf. e.g., Habel & Tappe, 1999). The content of a route direction might also be dependent of factors like the information offered, the time limit (Wahlster et al., 1998), and the salience of map objects and of the depicted real world objects (cf. footnote 8). As we already pointed out, for the dynamic presentation mode, the perceptual saliency conditions could be different than for the static presentation mode.

2.5 Spatial Chunking

We start this section with a short discussion of three features of route conceptualization which play a core role in our investigation of spatial chunking, namely landmarks, decision points and ordering information.

Landmarks. Additionally to a given street network salient geographical features are employed as external reference points often called landmarks. In route directions they function as adjustments between a built up representation and the actual spatial environment and are, moreover, of prime importance for learning and retrieving spatial information. They are general basic organizational features, cues within the route (Presson & Montello, 1988; Golledge, 1999). In our study we reduced the

⁹ A similar mechanism applies in the conceptualization of event-structures: Events adhere to a temporal precedence relation induced by their chronological order. Yet in verbalizing events, speakers construct hierarchical event structures and select either sub-ordinate or superordinate event knots for verbalization (cf. Habel & Tappe, 1999).

meaning of landmarks to identifiers of decision points, i.e. a landmark is associated with an intersection in the near vicinity, to allow for reference to the landmark instead of the intersection.

Decision Points. Decision points (DPs) are operationalized as any type of intersection where streets join (as opposed to non-decision points, which are locations along streets between intersections). In other words, at decision points it is necessary to make a decision since there are alternatives to continue, i.e., it is possible to change direction. When acquiring route knowledge more information is coded at intersections of paths, where choices are made, as opposed to between intersections. Decision points receive a lot of attention in route directions as they afford viewpoints to actual and potential navigation choices. Generally, speakers are aware of the complex environmental information they have to encode.

Ordering Information. As mentioned above, routes are curves, i.e., oriented linear objects (cf. Eschenbach et al., 1999). When reaching a decision point, the main question to decide is whether the instructed person has to go straight or has to turn. On the other hand, the instructor—i.e., who produces a verbal route description while perceiving a route map—has to detect in the stimulus route, which configurations along the route constitute decision points. With respect to a particular decision point, the orientation of a turn is the relevant information to communicate. We see turn-off constellations as sections of a path, which divide their surrounding into a left and a right half plan, induced by the orientation of the movement (cf. Schmidtke, Tschander, Eschenbach & Habel, in press). This property is valuable for a functional differentiation of route sides at decision points. They can clearly be discriminated by the value of the angles, which enclose them, one inside angle, being smaller than 180° and one outside angle, being larger. The side with the smaller angle is the functionally relevant side: Additional branching-off streets on the functionally relevant side directly influence the determinacy for decision-making both in navigation and in route descriptions. ‘Turn right’ is an unambiguous expression as long as there is only one possibility to turn right. In contrast to this, additional branching-off streets on the functionally irrelevant side may distort the internal spatial structure of the decision point but do not necessarily result in ambiguity or wrong decisions. As long as instructors let navigators know that they have to make a right turn at a given intersection the number of branches on the functionally irrelevant side are of minor importance.

In accordance with the fact that the linearization problem for route instructions does arise in a specific way (cf. 2.3), the question emerges how parts of the path are chunked and integrated into a route direction and if there are differences in chunking depending on the presentation mode. A ‘complete’ route direction would include every feature along the route. In the case study presented in section 3 we identify decision points and landmarks as being mayor features for spatial chunking. Decision points can be subdivided in two categories: DPs which afford a *directional change*, for short *DP+* and DPs *without a directional change*, abbreviated as *DP-*. Whereas a *DP-* is a good candidate to be chunked, the *DP+* are especially crucial for a route direction because they constitute *change points*. If the addressee misses a *DP+*, then there is the risk of going astray and loosing orientation. In the consequence, a *DP+*

should not be seen as ‘chunkable’ in the specific task of giving a route instruction, since this could result in losing information that is vital for conveying the route. We identify three ways for chunking the spatial information in between any two DP+.

The first possibility employs counting the DP- that are situated in between two DP+, or, alternatively, between an actual position of the addressee and the next DP+. We dub this strategy *numerical chunking*. It is evidenced by phrases like: *Turn right at the second intersection*.

The second possibility utilizes a non-ambiguous landmark for identifying the next crucial a DP+ and is thus called *landmark chunking* in the following. The employment of landmark chunking becomes apparent in phrases like: *Turn left at the post office*.

There is a third alternative—henceforward called *structure chunking*—that is based on a spatial structure being unique in a given local environment. Such a distinguished spatial configuration, like for example a T-intersection, can serve the same identifying function as a landmark. If the direction of traveling is such that the spatial structure appears to be canonically orientated (cf. figure 1b), the structure as such is easily employable for spatial chunking, resulting in utterances like *Turn right at the T-intersection*. A T-intersection is such a salient feature that it is recognizable, even if the direction of traveling does not result in it being canonically orientated, cf. fig. 1a. Although the intersection does not look like a T-intersection from the route perspective, our route instructors used utterances like, *turn right at the T-crossing* in analogous situations.

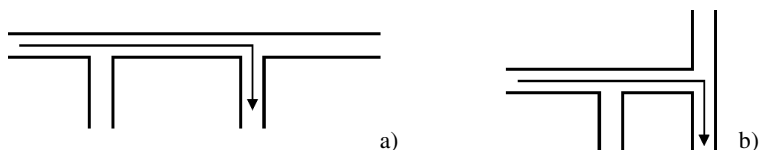


Fig. 1. The uniqueness of a spatial structure, i.e. employing the spatial structure as a landmark, dependent of the direction of traveling.

In all three cases of spatial chunking, the number of intermediate decision points or other route features is not specified a priori. It is sensible to assume, however, that the number of left-out-DP, i.e. the DP without directional change (DP-), is not arbitrary. A route direction like *Turn right at the 25th intersection* is unlikely to occur as it violates processability assumptions that the speaker implicitly applies. In other words, it is part of the addressee model that human agents are not primarily processing quantitative measures in spatial discourse.¹⁰

¹⁰ The maximal number of chunkable intersections is dependent on the spatial situation and is not in the focus of this research. The respective parameters in instructing a mobile artificial agent will be quite different from that of human agents.

3 A Case Study on Accompanying Route Directions from a Map

To shed light on the research questions raised in the previous sections we conducted a case study with a route presented in a map in two ways, statically and dynamically. With this distinction we aim at gaining insights in the processing of spatial information while producing accompanying route directions from an external representational medium. We are thus starting out from a *medium perspective* (what is the spatial structure actually depicted in the map?) and analyze the language data from a *procedural perspective* (which types of spatial structures are construed by the speakers during verbalization?). According to a long-standing tradition in the cognitive sciences, we use verbalizations as an empirical method to get access to otherwise hardly obtainable internal conceptualization processes. Specifically, we elicited accompanying route directions without hearer feedback. This has the advantage that we got longer discourses, where the structuring of the textual information partly reveals the presumable structure of the underlying internal representations on part of the speakers.

3.1. Material

Giving accompanying route directions from a veridical representational medium, i.e. a map with faithful information on angles and distances ensures that conceptual processes are not influenced by memory constraints, as might be the case for in advance route directions. The stimulus map¹¹ (see Fig. 2) was built on topographic data of the street network of a middle-sized town in Germany, slightly changed to fit the task in two ways: First, we added different kinds of landmarks which have proved in pre-tests to be easily recognizable. In Figure 2 we present the variant for the German verbalizers; the US-American participants received the same map with the same landmark locations albeit with US-American icons (e.g. McDonald, K-Mart). Second, we inserted a small number of additional branching-off streets in order to aggravate predictions about the route's continuation and thus to make spatial chunking more difficult in the on-line presentation mode. For the same reasons, we indicated only the route's origin in the map (by a visually salient green flag) but did not highlight the route's destination.

The route as depicted in Fig. 2 was presented either as a solid line, i.e. *static presentation mode*, or, as a moving dot, i.e. *dynamic presentation mode*. We chose the route according to the following criteria:

- The overall direction of the route is from right to left, i.e. against the usual reading/writing direction.
- The route is long enough to include a number of left and right turns.
- The route passes different kinds of intersections.
- It allows the participants to use different kinds of spatial chunking.

¹¹ The streets of the stimulus are built on the spatial relations of a topographic map, which means that they are veridical with respect to the spatial information that can be inferred from them, for example angles and distances. On the other hand, the graphic realization was simplified and certain features were left out.

received course credit for their participation. Two German and three US-American participants had to be excluded from the sample because their language output was hardly comprehensible (low voice quality).

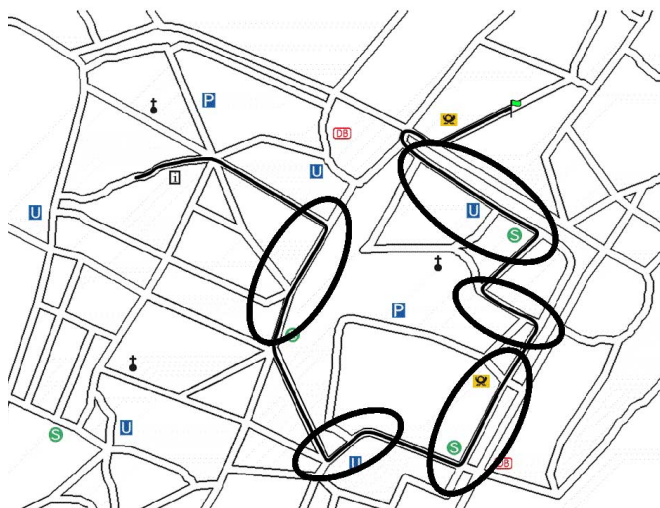


Fig. 3. Route segments that are situated between two DP+ and thus are candidates for spatial chunking.

3.3 Procedure

Participants were divided into two groups, a *dynamic condition group* and a *static condition group*. They were tested individually in an inter-individual design. Written instructions embedded the language production task into a communicative setting: First part (for both groups).

You are an employee at the central office of a modern messenger-service. There are plans to create the technical means to observe the messengers' movements on a screen and—for example in case of delay due to the traffic situation—to transmit them alternative routes by radio.

In order to practice, a training scenario has been developed, which we are going to demonstrate now.

Continuation of the scenario with alternations for the static/dynamic presentation of the route:

In this scenario you can see a line/a dot that is drawn into the map/ moves across the map and that suggests a path, which one of the messengers could take. The green flag marks the starting position. Please try to give the messenger a route instruction that is as precise as possible.¹²

¹² The static condition group was informed about the acoustic signals and their significance (cf. 3.1).

Additionally, participants were encouraged to ask questions and were instructed to watch carefully what happens and to simultaneously produce an accompanying route instruction that is suitable for reaching a destination at the end of the presented route. Subsequently, participants were asked to press an ‘O.K.’ button on the screen to start the *Flash* movie. They saw a countdown from 5 to 1, then the map appeared. The route’s origin (as marked by a little green flag, cf. 3.1) was at the same position as the count-down-numbers in order to avoid visual search.

The *dynamic condition group* received the map with a point moving through it. The verbalizers produced accompanying route instructions on the basis of the movements of the point, i.e. they began their route instruction as soon as the point appeared and they stopped shortly after it had reached its destination.

The *static condition group* was presented the same map. Instead of being presented a moving point, the route appeared as a solid line. Participants began their route instruction as soon as the map (with the route drawn into it) appeared and they stopped when their route instruction had reached the destination. None of the speakers ran out of time.

3.4 Predictions

We are interested in the effects of the presentation mode—static versus dynamic—on the processing of spatial information while speakers are producing accompanying route directions without hearer feedback from an external representational medium. More specifically, we focus on the spatial chunking of route segments and map features as is evidenced in the language data. Our predictions were the following:

Prediction 1-Visual accessibility influences spatial chunking

In the static presentation mode the route is drawn into the map as a bold black line. It is visually accessible throughout the verbalization task, which allows preplanning; i.e. the speakers’ attention may scan route’s continuation prior to actually verbalizing it. As compared to this, in the dynamic presentation mode the route’s continuation is not accessible to the speakers. Here they are giving the route instruction nearly simultaneously to the dot’s movement through the map. Thus, spatial chunking is discouraged. In the consequence static presentation should allow for more spatial chunking than dynamic presentation.

Prediction 2-Speakers avoid spatial chunking in accompanying route directions

In our setting speakers are producing accompanying route descriptions while they are exposed to a spatial environment, they do not know. Thus they can reduce cognitive processing costs in adhering to the local spatial structure at every moment in time and refrain from building up higher order spatial segments. They may think, that under such conditions, spatial chunking is prone to error and leads to misguiding their addressee. These effects should, again, be especially strong for the speakers in the dynamic presentation group who have reduced chances of pre-planning.

3.5 Scoring / Coding of Data

As discussed in section 2.5, we distinguish between different sub-types of spatial chunking during task-specific processing. Chunking is evidenced in the language

data, when decision points are not explicitly mentioned but are integrated into super-ordinate units; as a result elementary route segments are combined to form super-ordinate route segments. The stimulus route comprises five route segments that allow for spatial chunking (see Fig. 3) and are separated by decision points with directional change (DP+). This also holds for route segments CD and DE: Even though the intermediate intersection might not at first sight appear to be a DP+, it was univocally treated as such by our participants. Following this logic, we use the route segments encircled in Fig. 1 as data points, i.e. here we counted whether or not spatial chunking occurred. At each of these route segments one or more than one kind of chunking can be employed. More specifically: Numerical chunking can be used in all five route segments, landmark chunking is applicable in segments AB, CD and DE, whereas structure chunking is only available in segments BC and DE. This latter point is closely linked to the interaction with the external medium. In the stimulus map only T-intersections were unambiguously identifiable as compared to intersections with several branching-off streets. In the scoring procedure we accounted for the fact that not all types of spatial chunking can be realized in all route segments by weighting the scores accordingly.

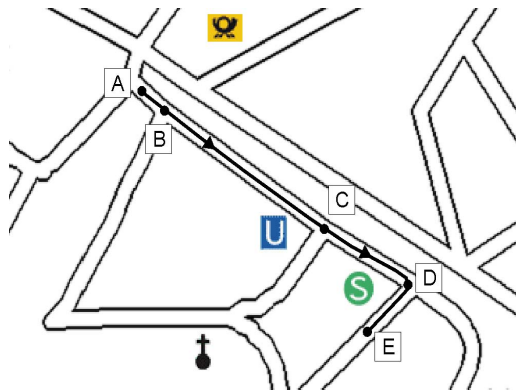


Fig. 4. Route segments (AB, BC, CD, DE) can be chunked to super-ordinate route segments in different ways. A route direction from the origin A to destination E can employ *numerical chunking*, i.e. ‘turn right at the third intersection’, or by *landmark chunking*: ‘turn right after the S-Bahn station’. The number of in-between decision points is unspecified.

The participants’ route descriptions were tape-recorded and transcribed in full. The transcripts were analyzed in terms of kind and quantity of chunked route segments. For the analysis of content, each transcript was divided into discrete utterances, and the authors rated relevant utterances according to the chunking types listed in Table 3. For each verbalization, we counted the number of complex nouns phrases that indicate a spatial chunking process. In cases where a speaker employed more than one kind of chunking in one phrase, we solely counted the first. An example like: *Turn right at the McDonalds, which is the second intersection* was coded as landmark chunking, i.e. *at the McDonalds*. An independent rater checked reliability of the analysis. Inter-rater agreement was 96% for chunking scores.

Table 3. Categories used to code utterances with examples.

Label	Category Name	Examples
LC	Landmark chunking	“turn left at the station”, “go straight after the post office”.
NC	Numerical chunking	“turn left at the third intersection”, “it’s the second street to the right”
SC	Structure chunking	“turn left at the T-junction”

In a first step we kept analyses for the German and the US-American verbalizers apart. Since we did not find significant differences between the two language groups and this paper does not focus on an intercultural comparison, we present the results in one body.

3.6 Results

In general, we found that spatial chunking figures in about 53,8 % of all cases across conditions. Thus our prediction (prediction 2) that speakers avoid spatial chunking in accompanying route directions was not fully met. Instead of adhering to the ordering of the spatial objects along the route in a strict sense, in half the cases they chose to form super-ordinate route segments. Thus our investigation underpins the finding that route instructors are striving to structure the to-be-conveyed spatial environment and to present relevant, non-redundant information. This holds despite the fact that they were producing accompanying route directions on-line.

Figure 5 depicts the mean values for the occurrence of the three kinds of chunking specified above for the two conditions—static and dynamic—, which are weighted according to the possibility to employ each type of chunking at each of the five route segments in question.

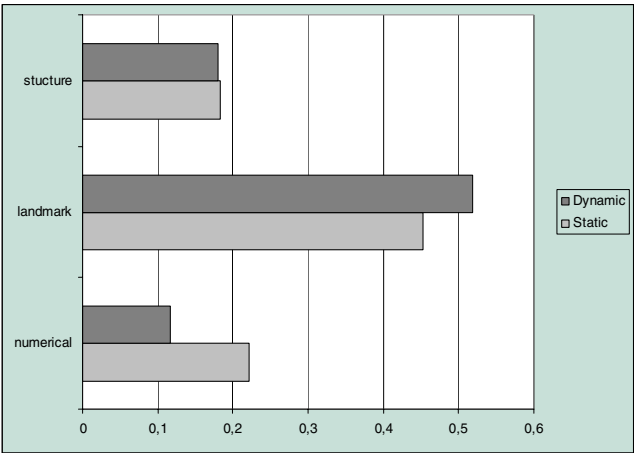


Fig. 5. Weighted mean values (numerical 5; landmark 3; structure 2) for three different kinds of chunking for the two conditions.

The results show the following pattern: Landmark chunking is the most common way to group primary route segments into secondary route segments underpinning the importance of landmarks for route directions from a procedural point of view. The importance of this finding is emphasized by the fact, that for landmark chunking we did not find significant differences between presentation modes. Almost the same pattern figures for structure chunking that was employed to a far lesser extent than landmark chunking: Presentation mode did not yield significant differences. Quite different from this pattern are the scores for numerical chunking: Presentation mode had a clear impact and we found a significant difference ($p=0.009$, ANOVA).

3.7 Discussion

As we see from the results of the case study, spatial chunking of elementary route segments is utilized as a helpful and adequate strategy in the production of route directions even in a setting where it adds to the cognitive processing load of the speakers. This holds especially for route directions that are processed during dynamic presentation mode: Here planning processes are aggravated because attention has to orient itself to the near vicinity of the moving dot in order to produce adequate guidance for the addressee. Even though speakers may visually scan the surroundings, the continuation of the route is not unerringly predictable. Thus a description of actions at every decision point—with or without directional change—seemed probable. However, even if verbalizers could in principle use all the information they had access to, they often chose not to do so. For example, instead of explicitly including every intersection along a straight part of the path into the route direction, people were likely to chunk segments together. These findings indicate that our second prediction (prediction 2, section 3.4), i.e. *speakers avoid spatial chunking in accompanying route directions*, was not met in an overall manner. What we found in the case study data was instead, that speakers attempted to use spatial chunking where they found it appropriate to the situation, even if it enhanced cognitive processing costs. This was the case in about half the cases overall.

Moreover, the results presented in section 3.5 indicate that the spatial chunking process especially utilizes landmarks and unambiguous spatial configurations—T-intersections in the stimulus material—in the same manner for both presentation modes. The unambiguous identifyability of T-Intersections seems to result from the interaction with the external graphical medium, i.e. the map. Whereas T-intersections present themselves as a salient feature largely independent of their orientation in a map, they might not function as such in route directions derived from memory of a real-world environment. This issue, however, awaits further investigation.

In contrast to landmark and structural chunking, we found significant differences between the presentation modes for numerical chunking, which is clearly favored in the static condition. These latter finding confirms our first prediction, i.e. *visual accessibility influences spatial chunking*. Whereas landmarks and salient spatial structures are visually accessible by quickly scanning the route and are obviously judged by the route instructors to be good cues for guidance, as they are assumed to be recognizable for the addressee of the route instruction independently of her or his current localization on the route, this is not the case for numerical chunking. First, in the dynamic presentation mode it might be difficult for the most part to keep track of the exact number of branching-off streets while producing the on-line instruction.

Second, the instructors have no feedback as to the current localization of the addressee. Therefore, they seem to take into consideration that a direction like *turn left at the third intersection* is to a great extent dependent on the progression of the addressee along the route and therefore prone to potential confusion.

Thus, despite the fact that chunking is an omnipresent characteristic of route directions overriding even the guidance of the presentation mode, there remain differences in the processing of static versus animated presentations.

4 General Discussion

Our research investigates the conceptualization of route segments into super-ordinate chunks during a route direction task, first, from a theoretical point of view and, second, in an explorative study. Theoretically the interaction between different representational formats—internal or external—requires a distinction of representational levels. In the case of user-map interaction it is a medium-perspective as such, a functional perspective and a procedural perspective. To elicit specific conceptual aspects of this interaction, i.e. the chunking of route segments, we collected data during a route direction task where the route was indicated either statically by a solid line, or dynamically by a moving dot. As it turned out from our theoretical considerations and from first results of the data analysis, the linearization process in language production is closely related to the chunking process in the case of verbal route directions generated from routes depicted in a map. Following Tversky and Lee (1999), who propose modality independent building blocks for routes, we assume that chunked spatial representations are not only crucial for language production but also for our conceptualization of routes and graphically conveyed route directions.

While verbalizing route instructions, speakers are thus not conformed with the problem of linearizing arbitrary spatial features. Rather they have to combine elements along a linear structure into sensible chunks. The finding that this occurs similarly across presentation modes is important to note. Even though the dynamic presentation strengthens the sequential character of the route, landmark and structure chunking occur in about the same amount of cases for both dynamic and static presentation. This indicates the existence of route direction principles that override specific characteristics of the two presentation modes to a certain degree. The observed effect may consequently be due to the fact that structuring route segments is part of our everyday life and as such a conventionalized skill that is employed even in demanding situations such as during dynamic presentation. On the other hand, the result that static presentation did not lead to a greater degree of landmark and structure chunking may in part be attributed to empirical findings made by e.g. Hegarty (1992). She found that observers of static diagrams mentally animate them in certain circumstances. If this also holds for statically conveyed routes, the difference between dynamic and static presentation would be diminished. This latter speculation invites subsequent empirical testing.

In addition to the similarities between presentation modes, we also found a significant difference for numerical chunking. This encourages further research in order to elucidate cognitive mechanisms entangled with either of the two presentation modes and to reveal effects of animation in distinguished situations. Furthermore, such research should explicate, in which contexts it is preferable to keep things simple and

rather employ static depictions. The latter point is emphasized by research on mental animation of static diagrams (cf. e.g. Hegarty, 1992 and Bogacz & Trafton (in press)). Here the question arises in which cases supplementary animation is prone to hinder diagram interpretation rather than enhance it. In the specific case of route directions further research might also reveal differences between static and dynamic presentation modes that can be attributed to theoretical considerations about different kinds of spatial knowledge, i.e. route and survey knowledge. Whereas route knowledge comprises procedural knowledge of a route as well as an egocentric perspective and thus might profit from dynamic presentation, survey knowledge fosters configurational aspects and a survey perspective, which might be favored by a static presentation mode. These aspects are beyond the scope of the current article and await further investigation.

Acknowledgments

This paper stems from collaborative research between the projects *Conceptualization processes in language production* (HA 1237-10) and *Aspect maps* (FR 806-8) both funded by the Deutsche Forschungsgemeinschaft (DFG). Our student assistants Nadine Jochims, Heidi Schmolck and Hartmut Obendorf were indispensable in a number of practical tasks. For invaluable help with the data collection we would like to thank Dan Montello. For comments we thank Carola Eschenbach, Lothar Knuf, Lars Kulik, and Paul Lee. We are also indebted to two anonymous reviewers for their helpful comments on an earlier draft of this paper.

References

- Agrawalla, M. (2001). Visualizing route maps. PhD thesis, Stanford University.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Augustine, M. & Coover, M. (1991). Simulation and information order as influences in the development of mental models. *SIGCHI Bulletin*, 23, 33–35.
- Barfield, W.L.R., Lim, R. & Rosenberg, C. (1990). Visual enhancements and the geometric field of view as factors in the design of three-dimensional perspective display. *Proceedings of the Human factors society—34th annual meeting, Orlando, Florida*. (pp. 1470–1473). Santa Monica, CA: Human Factors Society.
- Barkowsky, T., & Freksa, C. (1997). Cognitive requirements on making and interpreting maps. In S. Hirtle & A. Frank (Eds.), *Spatial information theory: A theoretical basis for GIS*. (pp. 347–361). Berlin: Springer.
- Berendt, B., Rauh, R., & Barkowsky, T. (1998). Spatial thinking with geographic maps: An empirical study. In H. Czap, P. Ohly, & S. Pribbenow (Eds.), *Herausforderungen an die Wissensorganisation: Visualisierung, multimediale Dokumente, Internetstrukturen* (pp. 63–73). Würzburg: ERGON-Verlag.
- Bell, S. (1995). Cartographic presentation as an aid to spatial knowledge acquisition in unknown environments. M. A. Thesis. Geography Department, UC Santa Barbara.
- Bogacz, S. & Trafton, G. (in press). Connecting internal and external representations: Spatial Transformations of Scientific Visualizations. *Foundations of Science*.
- Bollmann, J. (1993). Geo-Informationssysteme und kartographische Informationsverarbeitung. In B. Hornetz & D. Zimmer (eds.), *Beiträge zur Kultur- und Regionalgeographie. Festschrift für Ralph Jätzold*. (pp. 63–73). Trier: Universität Trier.

- Buhl, H.M., Katz, S., Schweizer, K. & Herrmann, T. (2000). Einflüsse des Wissenserwerbs auf die Linearisierung beim Sprechen über räumliche Anordnungen. *Zeitschrift für Experimentelle Psychologie*, 47, 17–33.
- Casakin, H., Barkowsky, T., Klippel, A., & Freksa, C. (2000). Schematic maps as wayfinding aids. In C. Freksa, W. Brauer, C. Habel, & K.F. Wender (Eds.), *Spatial Cognition II – Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications* (pp. 54–71). Berlin: Springer.
- Daniel, M.-P. & Denis, M. (1998). Spatial descriptions as navigational aids: A cognitive analysis of route directions. *Kognitionswissenschaft*, 7, 45–52.
- Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive*, 16, 409–458.
- Denis, M., Pazzaglia, F., Cornoldi, C. & Bertolo, L. (1999). Spatial discourse and navigation: An analysis of route directions in the city of Venice. *Applied Cognitive Psychology*, 13, 145–174.
- Eschenbach, C., Habel, C. & Kulik, L. (1999). Representing simple trajectories as oriented curves. In A. N. Kumar & I. Russell (eds.), *FLAIRS-99. Proceedings of the 12th International Florida AI Research Society Conference*. (pp. 431–436). Orlando, Florida.
- Freksa, C. (1999). Spatial aspects of task-specific wayfinding maps: A representation-specific perspective. In J. S. Gero & B. Tversky (eds.), *Proceedings of visual and spatial reasoning in design*. (pp. 15–32). University of Sydney: Key Centre of Design Computing and Cognition.
- Ghaëm, O., Mellet, E., Tzourio, N., Bricogne, S., Etard, O., Tirel, O., Beaudoin, V., Mazoyer, B., Berthoz, A., & Denis, M. (1998). Mental exploration of an environment learned from a map: A PET study. Fourth International Conference on Functional Mapping of the Human Brain, Montréal, Canada, 7–12 juin 1998. *NeuroImage*, 7, 115.
- Golledge, R.G. (1999). *Human wayfinding and cognitive maps*. In Golledge, R.G. (ed.), *Wayfinding behavior*. (pp. 5–45). John Hopkins University Press: Baltimore.
- Golledge, R.G.; Dougherty, V. & Bell, S. (1995). Acquiring spatial knowledge: Survey versus route-based knowledge in unfamiliar environments. *Annals of the Association of American Geographers*, 1, 134–158.
- Habel, C. (1988). Prozedurale Aspekte der Wegplanung und Wegbeschreibung. In H. Schnelle / G. Rickheit (Hrsg.): *Sprache in Mensch und Computer* (pp. 107–133). Westdeutscher Verlag: Opladen.
- Habel, C. & Tappe, H. (1999). Processes of segmentation and linearization in describing events. In R. Klabunde & C. v. Stutterheim (eds.), *Representations and processes in language production*. (pp. 117–152). Wiesbaden: Deutscher Universitätsverlag.
- Hegarty, M. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18(5), 1084–1102.
- Herrmann, T., Schweizer, K., Janzen, G., & Katz, S. (1998). Routen- und Überblickswissen – konzeptuelle Überlegungen. *Kognitionswissenschaft*, 7, 145–159.
- Herrmann, Th., Buhl, H.M. & Schweizer, K. (1995). Zur blickpunktbezogenen Wissensrepräsentation: Der Richtungseffekt. *Zeitschrift für Psychologie*, 203, 1–23.
- Hunt, E., & Waller, D. (1999). *Orientation and wayfinding: A review* (ONR technical report N00014-96-0380). Arlington, VA: Office of Naval Research.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- Jones, S. & Scaife, M. (2000). Animated diagrams: An investigation into the cognitive effects of using animation to illustrate dynamic processes. In M. Anderson, P. Cheng & V. Haarslev (eds.): *Theory and application of diagrams: First International Conference, Diagrams 2000, Edinburgh, Scotland* (pp. 231–244). Berlin: Springer.
- Kaiser, M., Proffitt, D., Whelan, S. and Hecht, H. (1992). Influence of animation on dynamical judgements. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 669–690.
- Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, MA.: Harvard UP.

- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. MIT Press: Cambridge, MA.
- Lovelace, K.L.; Hegarty, M. & Montello, D.R. (1999). Elements of good route directions in familiar and unfamiliar environments. In C. Freksa & D.M. Mark (eds.), *Spatial information theory. Cognitive and computational foundations of geographic information science*. (pp. 65–82). Berlin: Springer.
- Maaß, W. (1994). From visual perception to multimodal communication: Incremental route descriptions. *AI Review Journal*, 8, 159–174.
- Maaß, W.; Baus, J. & Paul, J. (1995). Visual grounding of route descriptions in dynamic environments. In Proceedings of the AAAI Fall Symposium on Computational Models for Integrating Language and Vision. MIT, Cambridge.
- MacEachren, A.M. (1995). *How maps work: Representation, visualization, and design*. New York: The Guilford Press.
- McNamara, T.; Hardy, J. K. & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 211–227.
- Morrison, J.B., Tversky, B., Betrancourt, M. (2000). Animation: Does it facilitate learning? In *AAAI Workshop on Smart Graphics*, Stanford, March 2000.
- Newcombe, N. S. & Huttenlocher, J. (2000). *Making space*. Cambridge, MA: MIT-Press.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Presson, C.C. & Montello, D.R. (1988). Points of reference in spatial cognition: Stalking elusive landmarks. *British Journal of Developmental Psychology*, 6, 378–381.
- Scaife, M. & Rogers, Y. (1996) External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, 45, 185–213.
- Schmidtke, H.R., Tschander, L., Eschenbach, C, Habel, C. (in print). Change of orientation, In E. van der Zee & J. Slack (eds.). *Representing direction in language and space*. Oxford: Oxford University Press.
- Schumacher, S., Wender, K.F., & Rothkegel, R. (2000). Influences of context on memory of routes. In C. Freksa, W. Brauer, C. Habel, & K.F. Wender (eds.), *Spatial Cognition II - Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*. (pp. 348–362). Berlin: Springer.
- Steven, A. & Coupe, P., (1978). Distortion in judged spatial relations. *Cognitive Psychology*, 10, 422–437.
- Tappe, H. (2000). Perspektivenwahl in Beschreibungen dynamischer und statischer Wegeskizzen. In C. Habel & C. von Stutterheim (eds.), *Räumliche Konzepte und sprachliche Strukturen*. (pp. 69–95). Tübingen: Max Niemeyer Verlag.
- Taylor, H. & Tversky, B. (1992). Descriptions and depictions of environments. *Memory and Cognition*, 20, 483–496.
- Thorndyke, P.W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14, 560–589.
- Tschander, L.B., Schmidtke, H.R., Eschenbach, C., Habel, C. & Kulik, L. (2002). A geometric agent following route instructions. In C. Freksa, W. Brauer, C. Habel & K. Wender (eds.), *Spatial Cognition III*. Berlin: Springer.
- Tversky B. (1993). Cognitive maps, cognitive collages and spatial mental models. In A. Frank & I. Campari (eds.) *Spatial information theory: A theoretical basis for GIS*. (pp. 14–24). Berlin: Springer.
- Tversky, B. & Lee, P.U. (1999). Pictorial and verbal tools for conveying routes. In C. Freksa, D.M. Mark (eds.), *Spatial information theory. Cognitive and computational foundations of geographic information science*. (pp. 51–64). Berlin: Springer
- Wahlster, W.; Blocher, A.; Baus, J.; Stopp, E. & Speiser, H. (1998). Ressourcenadaptive Objektkalisation: Sprachliche Raumbeschreibung unter Zeitdruck. In *Kognitionswissenschaft*, 7, 111–117.
- Wahlster, W.; Baus, J.; Kray, C. & Krüger, A. (2001). REAL: Ein ressourcenadaptierendes mobiles Navigationssystem, *Informatik Forschung und Entwicklung*, 16, 233–241.
- Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, 21, 179–217.
- Zhang, J. & Norman, D. A. (1994). Representation in distributed cognitive tasks. *Cognitive Science*, 18, 87–122.

Self-localization in Large-Scale Environments for the Bremen Autonomous Wheelchair

Axel Lankenau, Thomas Röfer, and Bernd Krieg-Brückner

Bremer Institut für Sichere Systeme, TZI, FB3, Universität Bremen,
Postfach 330440, 28334 Bremen, Germany.
`alone@tzi.de`, `roefer@tzi.de`, `bkb@tzi.de`

Abstract. This paper presents **RouteLoc**, a new approach for the absolute self-localization of mobile robots in structured large-scale environments. As experimental platform, the Bremen Autonomous Wheelchair “Rolland” is used on a 2,176m long journey across the campus of the Universität Bremen. **RouteLoc** poses only very low requirements with regard to sensor input, resources (memory, computing time), and a-priori knowledge. The approach is based on a hybrid topological-metric representation of the environment. It scales up very well, and is thus suitable for self-localization of service robots in large-scale environments. The evaluation of **RouteLoc** is done with a pure metric approach as reference method. It compares scan-matching results of laser range finder data with the position estimates of **RouteLoc** on a metric basis.

1 Introduction

1.1 Motivation

Future generations of service robots are going to be *mobile* in the first place. Both, in classical application areas such as the cleaning of large buildings or property surveillance, but especially in the context of rehabilitation robots, such as intelligent wheelchairs, mobility will be a major characteristic of these devices. After having shown that it is technically feasible to build these robots, additional requirements will become more and more important. Examples of such demands are the operability in common and unchanged environments, adaptability to user needs, and low material costs. To satisfy these requirements, methods have to be developed that solve the fundamental problems of service robot navigation accordingly. Apart from planning, the primary component for successful navigation is self-localization: a robot has to know where it is before it can plan a path to its goal.

Pursuing these considerations, a new self-localization approach was developed for the rehabilitation robot “Rolland” (see Fig. 1a and [12,21]) within the framework of the project *Bremen Autonomous Wheelchair*. The algorithm is called **RouteLoc** and requires only minimal sensor equipment (odometry and two sonar sensors), works in unchanged environments and provides a sufficient precision for a robust navigation in large building complexes and outdoor scenarios.

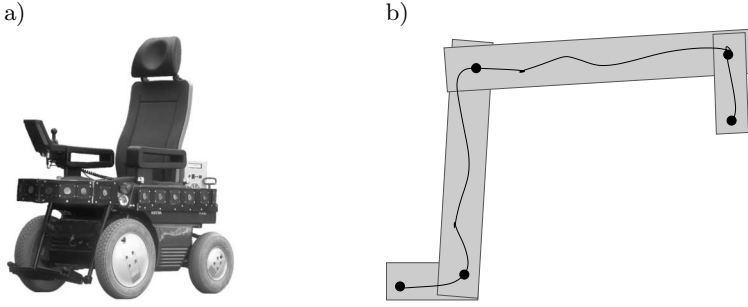


Fig. 1. a) Bremen Autonomous Wheelchair “Rolland”. b) Route generalization [18].

1.2 The Bremen Autonomous Wheelchair

The Bremen Autonomous Wheelchair “Rolland” (cf. Fig. 1a) is based on the commercial power wheelchair Genius 1.522 manufactured by the German company Meyra. The wheelchair is a non-holonomic vehicle that is driven by its front axle and steered by its rear axle. The human operator controls the system with a joystick. The wheelchair is equipped with a standard PC (Pentium III 600MHz, 128 MB RAM) for control and user-wheelchair interaction tasks, 27 sonar sensors, and a laser range sensor behind the seat. The SICK laser range finder has an opening angle of 180° toward the backside of the wheelchair and is able to deliver 361 distance measurements every 30 ms. The original Meyra wheelchair already provides two serial ports that allow to set target values for the speed and the steering angle as well as determining their actual values. Data acquired via this interface is used for dead reckoning. The odometry system based on these measurements is not very precise, i.e. it performs well in reckoning distances but it is weak in tracking angular changes. A modular hardware and software architecture based on the real-time operating system QNX allows for the adaptation to an individual user [22]. At the moment, the two main applications already implemented are the *Driving Assistant* and the *Route Assistant* [12].

2 Modeling Locomotion and Environment

Self-localization of robots is usually done by matching the robot’s *situation*, i.e. the current (and maybe also the past) sensor impressions and its locomotion, with a representation of its environment, e.g. a map. For a successful matching, it is indispensable that the models for both, the robot’s situation and the environment are comparable. The following two sections present the situation model and the environment model chosen for RouteLoc.

2.1 Situation Model

Röfer [18] introduces an incremental generalization of traveled tracks (see also [15]). The idea is to generalize the locomotion of the traveling robot during

runtime to an abstract route description. Such a description represents the route as a sequence of straight segments that intersect under certain angles. Since natural minor deviations occurring while traveling are abstracted away this way, the generalized description of the robot’s route from its starting point to its current location is an adequate situation model.

Specifying Abstract Route Descriptions. Fig. 1b shows the locomotion of the robot as recorded by its odometry system as a solid curved line. The corners recognized by the generalization algorithm are depicted as circles. The rectangular boxes represent the so-called acceptance areas: As long as the robot remains within such a region, it is assumed that the robot is still located in the same corridor. The width of the rectangular boxes is determined with the help of a histogram-based approach from the measurements of two sonar sensors mounted on the wheelchair’s left- and right-hand side chassis [18]. Note there may be other generalization algorithms that do not rely on external sensor input.

As a result, the generalization R of the route traveled so far is defined as a sequence of *corners* as follows:

$$R = \langle c_i \rangle, \text{ where } c_i = (\rho_i, l_i), i \in \{0, \dots, n\} \quad (1)$$

In contrast to the concept “corner_c” proposed by Eschenbach *et al.* [6], the length of the incoming segment of a corner is not considered here. In (1), ρ_i is the rotation angle between the incoming and the outgoing segment of a corner in a “local frame of reference”, i. e., ρ_i describes the relative change in orientation when passing corner c_i . As an example, consider the almost rectangular corner c_1 in the lower left part of Fig. 1b (c_0 is the “virtual” starting corner): ρ_1 is about 86° , because the robot has to turn about 86° to the left when changing corridors at corner c_1 . Note that ρ_0 is a “don’t care” value, i. e. only the outgoing segment of the first corner is considered, whereas the angle is ignored. The second parameter of a corner as specified in (1) is the length l_i of the outgoing segment.

Incremental Generalization of Route Descriptions. Since the situation of the robot has to be known while it travels, the route generalization must be carried out incrementally and in real-time. Röfer’s approach satisfies both requirements. Nevertheless, the incremental generalization has the drawback that it has to partially rely on uncertain knowledge: the distance l_n already traveled in the so far final segment as well as the angle ρ_n to the previous segment may change during runtime depending on the locomotion of the robot. The information about c_n is volatile and not fixed before a new final corner c_{n+1} is detected. This is illustrated in Fig. 2: The upper row of the figure shows three different snapshots of a single trajectory driven by the robot. The respective current location of the robot is indicated by the arrow. Even though this is only a sketch, it is reasonable to expect a similar odometry recording when the robot travels in a straight corridor, turns right after some time and turns left some time later. In the lower row, the corresponding generalizations are shown: In Fig. 2a, no corner

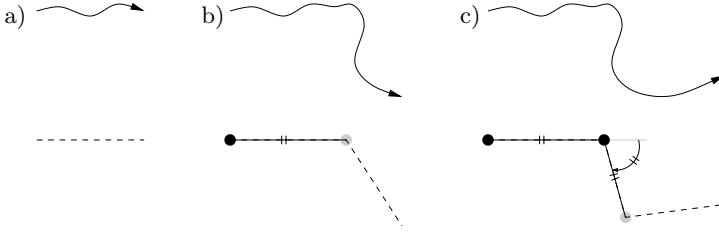


Fig. 2. Fixing of the penultimate corner during the incremental generalization.

has been detected so far, the traveled path completely fits into the imaginary corridor defined by the acceptance area of the segment depicted as a dashed line. In Fig. 2b, the robot has conducted a right turn and seems already to perform a new turn to the left. Nevertheless, it is only then that the robot leaves the acceptance area of the first segment. As a result, the generalization algorithm sets up a new—so far final—corner (indicated by the grey circle) and a new—also so far final—segment (indicated by the dashed line). Simultaneously, the parameters of the first corner c_0 (marked by the black circle) are fixed. Since it is the first corner, the angle is irrelevant; but the length of the outgoing segment is known now. In Fig. 2c, the robot has moved further and has left the acceptance area of the second route segment, resulting in the generation of another new segment. The generalization algorithm positions the third corner and fixes the parameters of c_1 : The rotation angle from the first to the second segment and the distance between c_1 and c_2 .

The abstraction resulting from this generalization method turns out to be very robust with regard to temporary obstacles and minor changes in the environment. Nevertheless, it is only helpful, if the routes are driven in a network of corridors or the like. Fortunately, almost all larger buildings such as hospitals, administration or office buildings consist of a network of hallways. In such environments, the presented algorithm works robustly.

2.2 Environment Model

In order to localize a robot within a representation of the environment such as a map, the model used for describing the current situation of the robot must be compatible to the model used for the description of the robot's environment, and it should be appropriate for the intended application scenario of the robot. Developing service robots, especially rehabilitation robots, usually means to develop low cost devices. Therefore, the equipment used for self-localization should be as sparse as possible. Nevertheless, mobile service robots such as cleaning robots, surveillance robots, and smart wheelchairs often have to cover a large operation space. That means that the self-localization approach must be able to work in large-scale environments such as complex buildings, university campuses or hospital areas. Especially in the context of rehabilitation robots the environment cannot easily be changed, e.g., by mounting artificial landmarks or beacons at

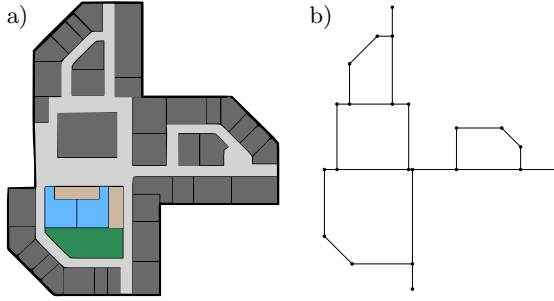


Fig. 3. a) Sketch of a floor. b) Corresponding route graph.

decision points, because they are often part of public buildings. Furthermore, environment changes are very expensive. As a consequence, an approach is needed that requires only minimal sensor equipment, works in unchanged environments that is able to operate reliably in large-scale scenarios.

Taking into account these aspects, a topological map that is enhanced with certain metric information appears to be an adequate representation of the environment in this context. Adapted from [30], such an environment model will be referred to as *route graph*. In the following, the nodes of a route graph correspond to decision points in the real world (or *places* as called by [30]): hallway corners, junctions or crossings. The edges of a route graph represent straight corridors that connect the decision points. In addition to the topological information, the route graph contains (geo-)metric data about the length of the corridors as well as about the rotation angles between the corridors. For example, Fig. 3a shows a sketch of the second floor of the MZH building of the Universität Bremen. The corresponding route graph is depicted in Fig. 3b. It consists of 22 nodes (decision points) and 25 edges (corridors) connecting them.

Since the route graph (environment model) has to be matched with route generalizations (situation model), it is advantageous *not* to implement the graph as a set of nodes that are connected by the edges, but as a set of so-called *junctions*:

Definition 1 (Junction). A junction j is a 5-tuple

$$j := (H, T, \gamma, o, I)$$

where H (“home” of the junction j) and T (“target” of j) are graph nodes that are connected by a straight corridor of length o . The set I consists of all incoming junctions j'_i that lead to j , i. e., $I = \{(H', H, \gamma', o', I')\}$. The function $\text{incomings}(j)$ selects the incoming junctions of j , i. e., $\text{incomings}(j) = I$. The signed angle γ is the rotation angle between the prolongation of an outgoing segment of some j'_i to the outgoing segment of junction j , i. e. it denotes by how many degrees it has to be turned to travel through j . For left turns, γ is positive; for right turns, γ is negative; $\gamma = 0$ means that j is a so-called “straight junction”, e. g. the T-bar of a T-junction (cf. Fig. 4).

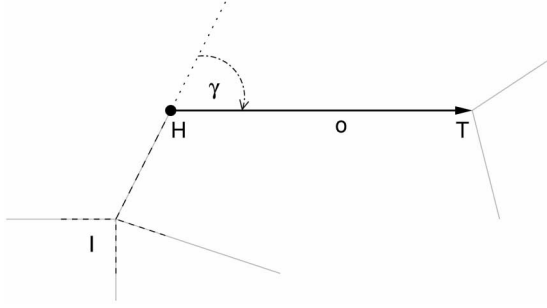


Fig. 4. Junction in a part of the route graph.

Note that outgoing segments of junctions are directed, i. e. junctions are one-way connections between route graph nodes. As shown in Sect. 3.1, the corners of a route generalization are compatible with the junctions of the route graph in that they can be matched and assigned with a real number representing a similarity measure.

Based on definition 1, a route graph G is the set of all junctions:

Definition 2 (Route Graph). *A route graph G is a set of all junctions that are connected:*

$$G = \{j = (H, T, \gamma, o, I) | \exists j' \in G : j' \neq j \wedge j' \in \text{incomings}(j)\}$$

While the representation of the environment as a route graph is formally similar to Voronoï diagrams as recently used, e.g., by Thrun [25], Zwynsvoorde *et al.* [31,32], and Choset [3], the localization approach presented here is not only applicable in sensory-rich (indoor) environments but also in pure outdoor or hybrid scenarios such as the campus example presented below. This is because the generalization of the robot's locomotion is used as reference information for the localization. Thus, **RouteLoc** does not have to rely on input from proximity sensors as it is necessary for the Voronoï diagram based approaches (a Voronoï diagram is defined on the basis of the sensor-perceived distance of the robot to objects in its environment).

In contrast to metric (grid-based) representations, the route graph is much easier to handle with respect to the required amount of computing time and memory. For example, the campus environment used for experiments in the results section (see Sect. 6 and Fig. 12a) is coded as a list of only 144 junctions (see Fig. 13b). The complexity of **RouteLoc** is linear in the number of junctions in the route graph. Therefore, it is important to note that covering a larger area with the route graph does not necessarily mean an increase in junctions. Instead, the critical question is, how many decision points are there in the environment.

3 RouteLoc: An Overview

This section is meant to explain how generalized route descriptions as situation model and a route graph as environment model are used for the absolute self-localization of a mobile robot in large-scale environments. First, a sketch of RouteLoc is presented to explain the basics of the algorithm. The simplifying assumptions made here for clarity purposes are then dropped in the detailed description of the algorithm in Sect. 4.

The basic idea of the self-localization approach is to match the incremental generalization of the currently traveled route with the route graph. This matching process provides a hypothesis about the robot's current position in its environment: RouteLoc ongoingly determines the hallway (represented by an edge in the route graph), in which the robot is most likely located at that very moment in time. Since the distance already traveled in the hallway is also known, an additional offset can be derived. As a result, the position of the robot within the hallway is found precise enough for most global navigation tasks. The precision is limited by about half of the width of the corridor the robot is located in, as is shown in Sect. 3.3.

3.1 Matching Route and Route Graph

Due to the dualism between a junction in the route graph and a corner in the generalized route, the chosen situation model and the environment model are compatible. Thus, self-localizing a robot by matching a generalized route with a route graph should in principle be straightforward. Nevertheless, there are some pitfalls that have to be paid attention for.

Since the algorithm has to deal with real data, there are almost no “perfect matches”. That means that even if the robot turned by exactly 90° at a crossing, the angle of this corner as calculated by the route generalization will almost certainly differ from 90° . This is mainly due to odometry errors. On the other hand, two corridors that meet in a perfect right angle in the route graph may well include an angle of only 89.75° in reality. These uncertainties have to be coped with adequately.

A second topic worth considering is the complexity of the matching process: At least in theory, a route can consist of arbitrarily many corners. Therefore, matching the whole generalized route with the route graph in each computation step is not feasible, because—at least in theory—this would require an arbitrarily long period of computing time. A solution to this problem is presented in the following subsections.

Within this section, it is assumed that every corner existing in reality is detected by the generalization algorithm *and* that every corner detected by the generalization algorithm is existing in reality. As mentioned earlier, this assumption is simplistic and unrealistic. Nevertheless, it is reasonable here in order to simplify the explanation of the basic structure of RouteLoc. The details of the algorithm are thoroughly discussed in Sect. 4.1.

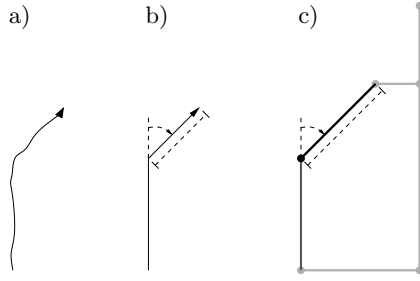


Fig. 5. Direct match of route corner and route graph junction. a) Odometry recorded. b) Corresponding route generalization. c) Matching junction in the route graph.

Direct Match of Route Corner and Graph Junction. If there are only two corners in the route, i. e. $R = \langle c_0, c_1 \rangle$ (the “don’t care” corner c_0 and the first “real” corner c_1), a *direct match* of c_1 and some junction j in the route graph is possible (cf. Fig. 5). As mentioned above, a binary decision of whether or not c_1 and j match is not adequate in this situation. Thus, a probabilistic similarity measure is introduced that describes the degree of similarity between the route corner and the junction as a real number between 0 and 1. For the route $R = \langle c_0, c_1 \rangle$ this value represents the probability that the robot is located in j .

The similarity measure m_d for the direct match of a route corner c with a route graph junction j is defined as

$$m_d(c, j) = s_l(c, j) \cdot s_\alpha(c, j) \quad (2)$$

In (2), the similarity s_l of the lengths of the outgoing segment of j and of the route segment of c , is defined as $s_l(c, j)$ with

$$s_l(c, j) = \text{sig} \left(1 - \frac{|l_c - d_j|}{d_j} \right) \quad (3)$$

In (3), l_c is the length of the outgoing route segment of c ; d_j is the length of the outgoing corridor of junction j . The longer the corridor, the larger the deviation may be for a constant similarity measure.

The similarity of the corresponding rotation angles, $s_\alpha(c, j)$, is defined as

$$s_\alpha(c, j) = \text{sig} \left(1 - \frac{\|\gamma_j - \rho_c\|}{\pi} \right) \quad (4)$$

In (4), γ_j is the rotation angle between the two segments of junction j , and ρ_c is the rotation angle of the final route corner c . Note that the result of this subtraction is always shifted into the interval $[0, \dots, \pi]$, as indicated by the $\|\dots\|$ notation. Please also note that these equations will be refined in the following in order to cover some special cases that will be introduced below.

In (3) and (4), the sigmoid function sig is used to map the deviations in length and in rotation angle into the intended range. The idea is to tolerate

small deviations with respect to the corridors' length or the angles, respectively, whereas large deviations lead to only small similarity values.

If the route R only comprises one corner (the “don't care corner”), i.e., $R = \langle c_0 \rangle$, the angle is ignored, because it is the initial rotation angle that has no meaning (cf. Sect. 2.1), thus $s_\alpha(c_0, j) = 1$. Therefore, the only remaining criterion for a direct match are the segments' lengths, thus $m_d(c, j) = s_l(c, j)$ in this case.

Induction Step. After having defined the direct matching for single-corner routes, the similarity measure has to be extended to longer routes. When a route $R = \langle c_0, \dots, c_n \rangle$ with $n > 1$ is to be matched with a junction j , it has to be found out, whether there is a direct match between corner c_n and junction j , and whether there is one between c_{n-1} and some j' with $j' \in \text{incomings}(j)$, and whether there is one between c_{n-2} and some j'' with $j'' \in \text{incomings}(j')$, and so on. If so, a sequence of junctions of the route graph is found the whole route R can be matched with.

Thus, the *matching quality* of a complete route R with respect to a specific route graph junction j is defined as follows:

Definition 3 (Matching Quality). *Given a route $R = \langle c_0, \dots, c_n \rangle$ with $n \geq 0$ and a junction j of the route graph G ($j \in G$), the matching quality $m(R, j)$ of R with respect to j is defined as*

$$m(R, j) = \max \left\{ \prod_{i=0}^n m_d(c_i, j_i) \mid \exists \langle j_0, \dots, j_n \rangle : j = j_n \wedge j_{k-1} \in \text{incomings}(j_k) \right\} \quad (5)$$

The definition states that every possible sequence of length $n + 1$ of route graph junctions is considered that fulfills two requirements: the final junction of the sequence must be j and the sequence must be “traversable”, i.e., the k -th junction in the sequence must be an incoming junction of the $(k + 1)$ -st junction of the sequence. As such a sequence consists of as many junctions as there are route corners, the matching quality can be determined by calculating the product of the direct matching qualities of the sequence junctions and the corresponding route corners. The overall matching quality of the route R and junction j is the maximum of all these products.

The number of such sequences grows exponentially with the length of the route and with the number of junctions in the route graph. Therefore, defining equation (5) is inadequate for a real-time capable localization approach. Fortunately, there is a workaround that dramatically reduces the complexity of calculating the matching quality: following the idea of the incremental route generalization, the matching quality can be defined inductively. In order to determine $m(\langle c_0, \dots, c_n \rangle, j)$, it is sufficient to know $m_d(c_n, j)$ and $m(\langle c_0, \dots, c_{n-1} \rangle, j')$

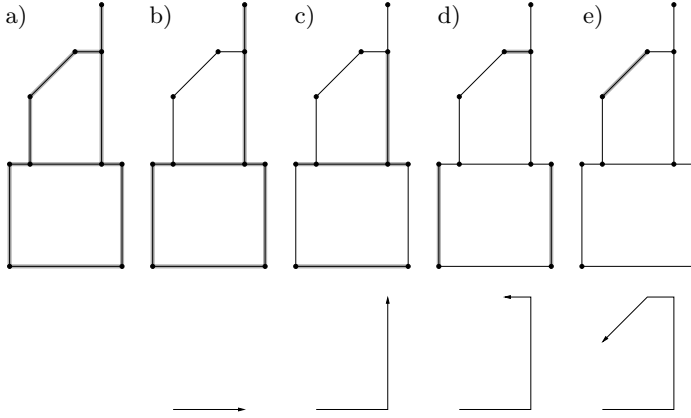


Fig. 6. Matching of route generalization and route graph.

with $j' \in \text{incomings}(j)$. As a result, the defining equation (5) can be refined to

$$\begin{aligned} m(\langle c_0 \rangle, j) &= m_d(c_0, j), & n = 0 \\ m(\langle c_0, \dots, c_n \rangle, j) &= m_d(c_n, j) \cdot \max_{j' \in \text{incomings}(j)} m(\langle c_0, \dots, c_{n-1} \rangle, j'), & n > 0 \end{aligned} \quad (6)$$

Calculating this recursion in every step is still impractical because it depends on the length of the route. Fortunately, the recursive function call can be avoided, if each junction is assigned with the probability value for having been in one of its incoming junctions before.

By applying definition 3 to the current route and every route graph junction, the junctions are assigned with a matching quality. The maximum of all the matching qualities provides a hypothesis which junction most likely hosts the robot. This junction is called *candidate junction* j_c for a route R .

$$j_c(R) = \operatorname{argmax}_{j \in G} \{m(R, j)\} \quad (7)$$

Figure 6 presents a step-by-step visualization of the localization process: In the initial situation, no information about the robot's potential location is available. Therefore, every junction in the graph can host the robot with the same likelihood. This is indicated by the edges underlined in grey in the route graph that is shown in the upper row of the figure. After the robot traveled some distance in a corridor (cf. Fig. 6b), three edges in the graph are identified in which the robot cannot be located. The route segment just traveled is longer than the corridors represented by these edges. After completing the first turn (90° to the left, see Fig. 6c), basically only three possibilities remain: Either the robot started in a corridor that is represented by one of the two facing edges depicted vertically in the lower part of the route graph or it started horizontally and its location is in the upper part of the graph afterwards. In Fig. 6d, another left turn yields no new information and thus no reduction of the possible robot locations. As

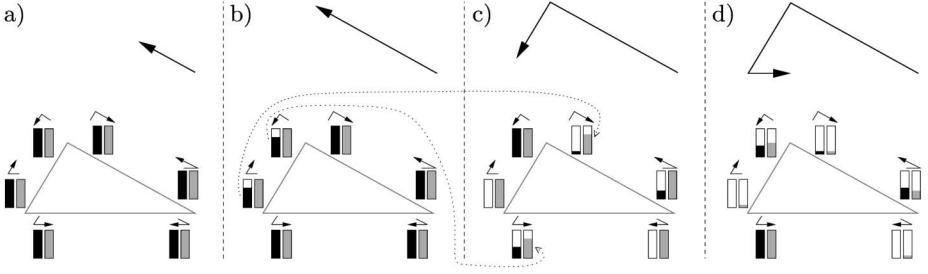


Fig. 7. Propagation of probabilities.

shown in Fig. 6e, the situation clarifies after the following turn: the location of the robot is determined by figuring out a unique candidate junction.

3.2 Propagation

In the previous subsection, it has been motivated that it is necessary to store the probability that the robot was in the incoming segment of a junction before detecting the last corner. But this information has to be transferred to other junctions when a corner in the route is detected by the generalization algorithm. After each corner detection, each junction is assigned with the maximum of the matching qualities of its incoming junctions, as discussed in Sect. 3.1:

$$m_{in}(j) = \max_{j' \in \text{incomings}(j)} m(\langle c_0, \dots, c_{n-1} \rangle, j') \quad (8)$$

Figure 7 shows four snapshots of a route traveled in a triangular environment. The upper part of each column shows the generalized trajectory as recorded by the robot. The arrow indicates the current position. The lower part of each snapshot depicts a route graph that consists of six junctions. Each junction is assigned with two probability values, depicted as partly filled columns. The left column (dark grey filled) indicates the direct matching quality of the final route corner with this junction. The right column (light grey filled) describes the probability of having been in the incoming segment of this junction before. A completely filled column stands for a 100% match, completely empty means something below 10% (but more than 0%). The arrows above the probability columns indicate the junction, e.g. the columns in the lower left corner of the route graph belong to the junction that leads from the left corridor to the lower corridor with a rotation angle of about 120°. From Fig. 7b to 7c, dotted arrows indicate the propagation.

3.3 Estimating the Robot's Position

Knowing the candidate junction and the offset already traveled in its outgoing segment enables RouteLoc to estimate a metric position of the form “The position

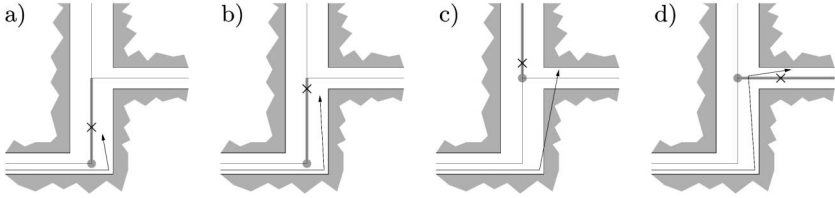


Fig. 8. “Generalization delay” when turning from one corridor to another.

is x cm in the corridor that leads from decision point A to decision point B .” One could argue that this metric information is superfluous for the user or for higher level navigation modules, because the corridors *between* the decision points are by nature free from decisions such as turning to a neighboring corridor. Thus, no detailed information about the robot’s location between the decision points should be required. Nevertheless, the metric information is indispensable for two reasons: First, not every location that is important for the robot’s task can be modeled as a decision point. Consider, e.g., some cupboard a wheelchair driver has to visit in a corridor. Second, when traveling autonomously, the robot often has to start actions or local maneuvers in time, i.e. they have to be initiated at a certain place in the corridor, maybe well before the relevant decision point can be perceived by the robot. This would be impossible without the metric information.

The rest of this section discusses some aspects that are relevant for a successful position estimate.

Ambiguous Environment. In some situations, the structure of the environment could turn out to be inadequate to this route-localization approach in its current version. In a square environment, for instance, the algorithm will fail, because every junction remains equally likely to host the robot even if the robot moves through the corridors. This problem of perceiving different places as if they were the same is commonly referred to as *perceptual aliasing*. When traveling in the square environment, four position estimates that are equally likely would be favored, no decision for a specific corridor would be possible.

Similarly, in a straight corridor, the algorithm is almost lost, because it has no means to infer where the robot started. Nevertheless, the longer the robot moves along the corridor, the less estimates are valid, simply due to the length of the trajectory already traveled. But even, if the robot traveled a straight route segment of about the corridor’s length, the algorithm still would generate two hypotheses about the robot’s position, one at each end of the corridor.

“Generalization Delay” when Changing Corridors. Due to the nature of the generalization algorithm, there exists a certain delay before the change of corridors can be detected. For example, in Fig. 8a, the generalization (depicted as a thin black line; the arrow indicates the current position) of the traveled

route is correctly matched with the route graph. The highlighted junction is the candidate junction, resulting in a position estimate which is indicated by the cross. The estimated position differs only slightly from the real position (cf. the paragraph on precision below). In Fig. 8b, the robot almost reached the T-junction. The localization is still correct. In Fig. 8c, the robot already changed corridors by taking the junction to the right. But the generalization algorithm has not yet been able to detect this, because it still can construct an “acceptance area” for the current robot position within the same corridor as before. Therefore, it assumes that the robot passed the T-junction and estimates the robot’s position to be in the junction that forms a straight prolongation to the former one. It is not until the robot has traveled some more distance before the generalization algorithm detects the corner (see Fig. 8d). Then, the position estimate is immediately corrected and a precise hypothesis is set up.

Precision of the Position Estimate. Because of the modeling of the environment and the robot’s locomotion, the algorithm is rather insensitive to odometry errors (see Fig. 12b). The offsets normally represent only short distances that result from accumulating straight movements, and almost no rotational motion which often causes dead reckoning errors. Nevertheless, the precision of the algorithm is limited to half the width of the current corridor at right angles to the robot’s driving direction and half the width of the previous corridor in the robot’s driving direction (see Fig. 9). The error could be even bigger, if the route graph is not correctly embedded in the center of the corridors, as it should be. Note that errors do not accumulate across junctions, but within longer junctions odometry errors may become significant.

The precision does explicitly *not* depend on the length of the traveled route, as every matching of a route corner to a graph junction once again limits the error. Nevertheless, the quality of the position estimate depends on the “quality” of the environment. The results of the experiments presented in Sect. 6 confirm this point of view.

4 Inside RouteLoc: A Deeper Insight

Section 3 uses the unrealistic assumption that the route generalization algorithm creates a new corner for every decision point (junction) the robot passes, and—vice versa—that every generated corner has its counterpart in the route graph and in the real world. This is too optimistic, as is shown below. Section 4.1 copes with this problem and presents a general solution that requires no restrictive assumptions.

Right at the beginning of a robot journey, a few special cases have to be paid attention to: If the robot did not start its travel at a decision point but within a corridor, the standard matching process as described above does not work as fast as it could. Furthermore, a route with no real corner detected so far requires some special attention during the matching process. This is discussed in Sect. 4.2.

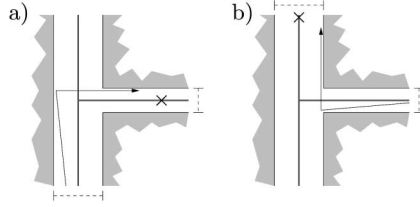


Fig. 9. Precision of the position estimate. a) Entering a narrow corridor from a wide one. b) Vice versa.

Another assumption made in Sect. 3 is that the robot can only change its general driving direction at decision points. This is a straightforward inference from the definition of decision points (junctions) and corridors connecting these decision points. But, there is a decision the robot can make anywhere, not only at decision points: turning around. Since the route graph junctions are directed, such a turning maneuver implies that the robot leaves the current junction. But unfortunately, it does not end in another junction represented in the route graph, because such turning junctions are not available in the route graph. Section 4.3 describes the handling of turning around within corridors.

4.1 On Phantom Corners and Missed Junctions

While the robot travels, the self-localization algorithm is expected to ongoingly present a hypothesis about the robot's current position. This hypothesis is to be updated in regular intervals. In the experiments presented in the results section 6, an update interval of 20cm travel distance has been used. In every update step, the route generalization algorithm checks whether a new corridor has been entered and updates the route description accordingly. Afterwards, the matching process is carried out which leads to a position estimate, as discussed in Sect. 3.3.

In every update step, four different situations can occur with respect to detected or undetected route corners, and to existing or not existing junctions in the route graph:

1. There is no junction in reality and the generalization algorithm correctly detects no route corner (see Fig. 10a). This is the normal case because most of the time the robot moves through corridors.
2. There is a junction in reality and the generalization algorithm correctly detects a corresponding route corner (see Fig. 10b). This was the assumption in the previous section.
3. There is no junction in reality even though the generalization algorithm detects a route corner, a so-called *phantom corner* (see Fig. 10c). Unfortunately, this case is not that rare, because due to odometry drift, long corridors are often generalized to more than one segment.
4. There is a junction in reality but the route generalization algorithm does not detect a corresponding route corner (see Fig. 10d). This is the problem

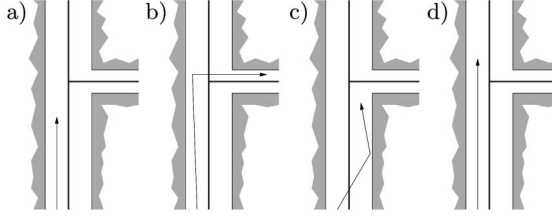


Fig. 10. Special cases handled by RouteLoc.

of *missed junctions* which is not a flaw of the route generalization algorithm but a result of the Spartan sensor use of the approach. Nevertheless, the self-localization algorithm is able to handle it.

The correct handling of these four situations is fundamental for the algorithm. They are discussed in the following sections.

There Is No Junction and No Corner Is Detected. In Fig. 10a, the standard situation is illustrated: the robot moves within a corridor, no junction in its surroundings, and the route generalization algorithm correctly infers that the robot did not change corridors, but still travels in the same corridor as one step before. In this case, the matching process can be carried out as described in Sect. 3. There is only one restriction: the definition of the similarity measure in (3) assumes that the final length of the route segment to be matched with the junction's outgoing segment is already known. As mentioned above, this is not the case for the currently final segment of the route traveled so far. Therefore, the calculation of the similarity measure $s_l(c, j)$ for the lengths of the final route corner c and a junction j has to be changed in this case to

$$s_l(c, j) = \begin{cases} 1, & l_c \leq d_j \wedge c = c_n \\ \text{sig} \left(1 - \frac{l_c - d_j}{d_j} \right), & \text{otherwise} \end{cases} \quad (9)$$

In (9), l_c is the length of the route segment of corner c ; d_j is the length of the outgoing corridor of junction j . In contrast to the original definition in (3), the similarity is set to 100% not only if the lengths are equal, but also if the final route segment is shorter than the junction segment. This is no surprise, as it is a preliminary match and the currently available information about the final route segment indicates that it matches the route graph junction. Only if l_c happens to be larger than d_j , the similarity measure drops below 100%. Note that (9) replaces (3) as definition of the similarity measure with respect to the segments' lengths.

As long as no corner is detected, there is no need for propagating the probabilities to adjacent junctions. Thus, the similarity values for each junction are only adapted to the current route generalization. Nevertheless, the case of missed junctions has to be kept in mind (see below).

There Is a Junction and a Corner Is Detected. In some situations, the route generalization algorithm detects corners in the route, as shown in Fig. 10b. If there exists a corresponding junction in the route graph, the matching as described in Sect. 3 will be successful. Note that detecting a new corner in the route fixes the then penultimate corner in its angle and length components. Therefore, the matching is a three-step process in this case: first, the new penultimate corner is matched according to the rules described in Sect. 3.1 and the similarity measure just defined in (9). Second, the probabilities are propagated to the adjacent junctions as discussed in Sect. 3.2. And third, the new final corner is matched as a non-fixed corner according to (9).

There Is No Junction, but a Corner Is Detected. Unfortunately, this case is not as rare as one could expect. As depicted in Fig. 10c, the motion track as recorded by the robot’s odometry can significantly deviate from a straight line even if the robot drives in a straight corridor. Especially in very long corridors, the odometry tends to be inaccurate. As an example, consider Fig. 12b that depicts the generalized motion track that has been recorded during experiments on the campus of the Universität Bremen. In the upper left part of the figure, the main boulevard of the campus, which is straight and about 300m long, is partitioned into several segments. This is because the odometry recorded the straight boulevard as a crescent-shaped curve. The erroneously detected “phantom corners” between the segments are a problem for the self-localization algorithm because the probability values have to be propagated through the graph after every route corner detection (see the section on propagation 3.2). If, however, such a detected route corner is a phantom corner, the propagation will be an error.

Therefore, when detecting a corner, the self-localization algorithm has to decide whether it is a corner with a corresponding junction in the route graph, or whether it is a “phantom corner” that results from bad odometry data. As if this were not enough, this decision cannot be made until the information about the route corner is fix. That means, the decision of whether or not a corner is believed to be either real or phantom can only be made with respect to the penultimate, already fixed, corner in the generalized route.

These considerations suggest to pursue two instead of one hypotheses for each junction (see Fig. 11): The first describes how probable it is that the robot is in the outgoing segment of the junction and has been in the incoming segment before the final corner has been detected (i. e., the final route corner is *real*; see Fig. 11c). The second hypothesis describes the probability that the robot is in the outgoing segment of the junction and has already been there before the final corner has been detected (i. e., the final corner is *phantom*; see Fig. 11d).

As a result, two similarity measures for the two hypotheses have to be defined: The similarity measure that assumes the final route corner to be a *real* corner is identical to m_d as defined in (2). It is renamed to m_d^r here. The similarity measure that assumes the final route corner to be a *phantom* corner is called

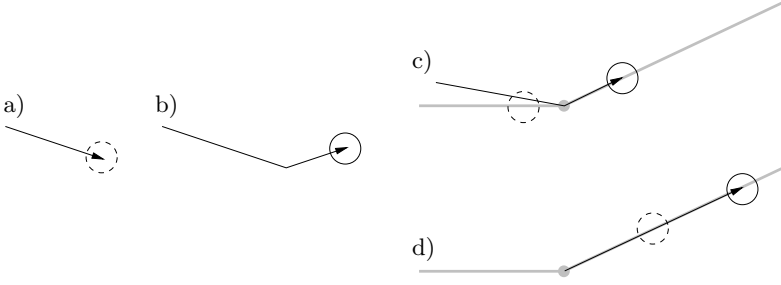


Fig. 11. *Real and Phantom route corners.* a) Generalized route before detection of the corner. b) After detection. c) *Real corner.* d) *Phantom corner.*

m_d^p . It uses (9) as measure for the similarity of the segments' lengths, but a different definition s_α^p of the rotation angle similarity:

$$s_\alpha^p(c, j) = \text{sig} \left(1 - \frac{\rho_c}{\pi} \right) \quad (10)$$

In (10), the rotation angle ρ_c of the route corner is compared to 0° , instead of to the junction angle as in (4). As a result, the matching probability is close to 100% for very small angles (i.e., detected route corners with a small angle are likely to be phantom corners) and low for significant angles (i.e., detected route corners with an angle of, say, 90° are expected to be real corners with high probability).

The two hypotheses are always considered in parallel, i.e., there are two probabilities for a junction to host the robot: One of them assumes that the final route corner is a real corner, which means that the robot has been in the incoming segment of the junction before the corner has been detected. The other one assumes that the final corner is a phantom corner, which means that the robot has already been in the outgoing segment of the junction before the corner has been detected. As a result, there also exist two matching qualities $m^r(R, j)$ (assuming final corner of R is real) and $m^p(R, j)$ (assuming the final corner of R to be phantom).

When a new final corner is detected in the route, the propagation process copies the superior hypothesis to the adjacent junction. At that time, a decision can be made about whether the real or the phantom probability is the “correct” one, because the corner is fixed in length and rotation angle.

The overall probability of the junction (i.e. the matching quality) is then calculated as the maximum of both hypotheses:

$$m(R, j) = \max\{m^r(R, j), m^p(R, j)\} \quad (11)$$

There Is a Junction, but No Corner Is Detected. It is possible that a corner existing in reality has been passed and has not (yet) been detected by

the generalization algorithm. As a consequence, the resulting change of corridors is not recognized (*missed junction*). Usually, this cannot be blamed on the generalization but on the fact that—based only on the locomotion data—one cannot distinguish traveling in a straight corridor with no junctions or crossings from traveling in a straight corridor passing several T-junctions. Therefore, the self-localization algorithm has to solve this problem. In every step, it is checked, whether the outgoing segment of the final route corner c_n is longer than the outgoing segment of the currently considered route graph junction j . If so, it is likely that this route segment is an overlap from a previous junction that leads to j . Note that not only straight predecessors of j (i.e. those that form a 0° angle with j) have to be considered here. Every incoming segment of j could have “hosted” the initial part of the route segment of corner c_n . Especially in long corridors with lots of crossings, it often happens that these overlaps stretch over more than one junction. ??).

Due to these considerations, it is always calculated how far the final route segment extends into the outgoing segment of the currently considered junction. This may significantly differ from the length of the final route segment. That is why it is a simplification to use the length l_c of the route segment in (3). Instead, in all equations for the similarity measure ((3), (9)), the distance l_c^+ already traveled in the segment has to be used instead of the length of the so far final route segment l_c (cf. Sect. 4.4).

4.2 Initial Phase Specialities

After solving the phantom corner and missed junction problems in Sect. 4.1, there are two special cases with respect to the early phases of a robot journey that are to be covered by the algorithm, but have not been addressed yet:

- Matching a route $R = \langle c_o \rangle$ that comprises only the initial corner with the route graph
- Starting the robot’s journey not at a decision point but somewhere in the middle of a corridor.

These two topics are discussed in the following two paragraphs.

Before the First Corner Was Detected. As discussed in Sect. 2.1, the rotation angle of the initial route corner c_0 is special in that it is a “don’t care” value. Even stronger, it may never be used during the matching process, because it has no meaning: it describes the rotation angle between the first route segment and an imaginary but not existing “zeroth” route segment. Therefore, the matching process has to be carried out slightly different as long as no real route corner has been detected. The implementation of this requirement is straightforwardly achieved by a further extension to the similarity measure calculation previously shown in (3) and refined in (9). The equation that includes the “before the first

corner” case looks as follows for the assumption that c_n is a real corner:

$$s_\alpha^r(c, j) = \begin{cases} 1, & c = c_0 \\ \text{sig}\left(1 - \frac{\|\gamma_j - \rho_c\|}{\pi}\right), & \text{otherwise} \end{cases} \quad (12)$$

and for the assumption that c_n is phantom:

$$s_\alpha^p(c, j) = \begin{cases} 1, & c = c_0 \\ \text{sig}\left(1 - \frac{\rho_c}{\pi}\right), & \text{otherwise} \end{cases} \quad (13)$$

where c_0 is the initial corner of the route.

Starting in the Middle of a Corridor. The basic idea of the whole approach is that detected route corners can be identified with certain junctions in the route graph. Then, the similarity measures deliver an adequate means to decide about the matching quality. However, at the very beginning of a robot journey, a situation may occur, where the robot does not start at a place in the real world that is represented by a route graph node. Instead, the starting position could be located somewhere in a corridor in the middle between two decision points. If the robot reached the first adjacent junction, detected a corner, and matched the route with the graph, the length of the driven segment would be significantly too short in comparison with the junction’s outgoing segment (because the robot started in the middle). Nevertheless, the route segment perfectly fits into the route graph. Thus, for the first route segment, it must be allowed that it is shorter than the junction’s outgoing segment without loss of matching quality. Once again, the equations for the similarity measures are refined to:

$$s_l^r(c, j) = \begin{cases} 1, & l_c^+ \leq d_j \wedge c \in \{c_0, c_n\} \\ \text{sig}\left(1 - \frac{l_c^+ - d_j}{d_j}\right), & \text{otherwise} \end{cases} \quad (14)$$

$$s_l^p(c, j) = \begin{cases} 1, & l_c^+ \leq d_j \wedge c \in \{c_1, c_n\} \\ \text{sig}\left(1 - \frac{l_c^+ - d_j}{d_j}\right), & \text{otherwise} \end{cases} \quad (15)$$

4.3 Turning Around within a Corridor

Nonholonomic vehicles such as the Bremen Autonomous Wheelchair “Rolland” are not able to move in arbitrary directions but they are restricted to bias bearings such as forwards and backwards instead. As a consequence, nonholonomic robots are not able to turn on the spot without shunting. Especially for the wheelchair, there are some corridors that are too narrow to turn at all. Therefore, it is fundamental to know the orientation of the wheelchair within a corridor. This is solved by modeling the corridors as one-way junctions, where the orientation is inherently known (see Sect. 2.2 on route graphs). If the robot turns around in a corridor, it leaves its current junction. But—by definition—leaving a junction means to enter another junction. Unfortunately, there are no junctions in the route graph that connect the two directions of a corridor.

An additional problem is that a turning maneuver can be carried out at any position within the hallway. In contrast to that, leaving the corridor is only possible at junctions.

To overcome these problems in order to be able to handle turns, the set of junctions that initially form the route graph G is extended by so-called “turn-junctions” at program start as shown:

$$G' = G \cup \left\{ (H, T, \pi, |\overline{HT}|, I) \mid \begin{array}{l} H, T \in N, I \subseteq G, \forall i \in I : i = (T, H, \gamma_i, |\overline{TH}|, I_i) \end{array} \right\} \quad (16)$$

In (16), for each junction j_i in the initial route graph G , all turn-junctions that can be generated for j_i are added to G . As an example, consider the route graph depicted in Fig. 13b that is used for the experiments presented in Sect. 6. The 144 junctions of this route graph require an additional set of 102 turn-junctions. The upper bound of the number of required turn-junctions for a route graph with n “real” junctions is $2n$. In typical environments, however, it often happens that two or more junctions share one turn-junction, e.g. junctions cdh and kdh in Fig. 13b both need the turn-junction dhd . The incoming and the outgoing segment of these turn-junctions represent the same hallway (forwards and backwards direction) and have a rotation angle of 180° . After having generated the turn-junctions at program start, they are dealt with as if they were “normal” junctions in the sequel. The only exception is that the deviation of the length is ignored when calculating the matching quality of a generalized route corner with such a turn-junction (undershooting is granted for turn-junctions).

4.4 Similarity Measures (Final Revision)

This section recapitulates the defining equations for the similarity measures including all special cases:

$$m_d^r(c, j) = s_l^r(c, j) \cdot s_\alpha^r(c, j) \quad (17)$$

$$m_d^p(c, j) = s_l^p(c, j) \cdot s_\alpha^p(c, j) \quad (18)$$

$$s_l^r(c, j) = \begin{cases} 1, & l_c^+ \leq d_j \wedge (c \in \{c_0, c_n\} \vee \text{isTurn}(j)) \\ \text{sig}\left(1 - \frac{l_c^+ - d_j}{d_j}\right), & \text{otherwise} \end{cases} \quad (19)$$

$$s_l^p(c, j) = \begin{cases} 1, & l_c^+ \leq d_j \wedge (c \in \{c_1, c_n\} \vee \text{isTurn}(j)) \\ \text{sig}\left(1 - \frac{l_c^+ - d_j}{d_j}\right), & \text{otherwise} \end{cases} \quad (20)$$

$$s_\alpha^r(c, j) = \begin{cases} 1, & c = c_0 \\ \text{sig}\left(1 - \frac{\|\gamma_j - \rho_c\|}{\pi}\right), & \text{otherwise} \end{cases} \quad (21)$$

$$s_\alpha^p(c, j) = \begin{cases} 1, & c = c_0 \\ \text{sig}\left(1 - \frac{\rho_c}{\pi}\right), & \text{otherwise} \end{cases} \quad (22)$$

5 Related Work

The following subsection gives a brief overview about mobile robot self-localization. In Sect. 5.2, **RouteLoc** is compared to prominent approaches and set in relation to Markov localization methods.

5.1 Self-localization Techniques

There are two basic principles for the self-localization of mobile robots [1]: *Relative* approaches need to know at least roughly where the robot started and are subsequently able to track its locomotion. At any point in time, they know the relative movement of the robot with respect to its initial position, and can calculate the robot's current position in the environment. It has to be ensured that the localization does not lose track, because there is no way to recover from a failure for these approaches. Modern relative self-localization methods make often use of laser range finders. They determine the robot's locomotion by matching consecutive laser-scans and deriving their mutual shift. Gutmann and Nebel [8,9] use direct correlations in their *LineMatch* algorithm, Mojaev and Zell [14] employ a grid map as “short term memory”, and Röfer [19] accumulates histograms as basic data structure for the correlation process.

On the other hand, *absolute* self-localization approaches are able to find the robot in a given map without having any a-priori knowledge about its initial position. Even more difficult, they solve the “kidnapped robot problem” [5], where—during runtime—the robot is deported to a different place without being notified. From there, it has to (re-)localize itself. That means, the robot has to deliberately “unlearn” acquired knowledge.

The absolute approaches are more powerful than the relative ones and superior in terms of fault tolerance and robustness. They try to match the current situation of the robot—defined by its locomotion and the sensor impressions—with a given representation of the environment, e. g. a metric map. As this problem is intractable in general, probabilistic approaches have been proposed as a heuristics. The idea is to pose a hypothesis about the current position of the robot in a model of the world from which its location in the real world can be inferred. A distribution function that assigns a certain probability to every possible position of the robot is adapted stepwise. The adaptation depends on the performed locomotion and the sensor impressions. Due to the lack of a closed expression for the distribution function, it has to be approximated. One appropriate model is provided by grid-based Markov-localization approaches that have been examined for some time: they either use sonar sensors [4] or laser range finders [2] to create a probability grid. As a result, a hypothesis about the current position of the robot can be inferred from that grid. Recently, so-called Monte-Carlo-localization approaches have become very popular. They use particle filters to approximate the distribution function [7,26]. As a consequence, the complexity of the localization task is significantly reduced. Nevertheless, it is not yet known how well these approaches scale up to larger environments.

Apart from these purely metric representations of the environment, Kuipers *et al.* propose the integration of metric and topological concepts with their “spatial semantic hierarchy” [11]. The idea is pursued by Simmons and Koenig [24] and Nourbakhsh *et al.* [16] by augmenting topological maps with metric information. The resulting self-localization methods also work probabilistically on the basis of the odometry and a local model of the environment perceived with the sensors. A very recent approach by Tomatis *et al.* combines map-building and self-localization [28]. They employ a 360° laser range finder and extract features such as corners and openings which are used to navigate in a global topological map. In addition, the laser-scans are searched for line structures (walls, cupboards, etc.) which build the basic data structure for several local metric maps (one for each node of the topological map).

5.2 Comparison between RouteLoc and Prominent Approaches

A number of prominent self-localization algorithms use the Markov localization approach, some of them with topological representations of the environment [24,16,28], others with metric maps [2,7,26]. In the robotics community, it is referred to as “Markov localization” if the algorithm somehow exploits the so-called *Markov assumption* [23]. It states that the outcome of a state transition may only depend on the current state and the chosen action. The outcome does explicitly not depend on previous states or actions.

RouteLoc is no pure Markov localization: while the matching and propagation process as presented in Sect. 3 satisfies the Markov assumption, the necessary handling of the missed junctions and phantom corners violates it. Apart from the “Markov or not” question, RouteLoc differs from other localization approaches with respect to some aspects that are gathered in table 1. As reference algorithms the topological-metric approach used for the office delivery robot Xavier by Simmons and Koenig [24] and the Mixture-MCL algorithm (an improved version of the common Monte Carlo Localization approaches) by Thrun *et al.* [27] are chosen.

6 Results

In order to evaluate the performance of an approach for the global self-localization of a mobile robot, a reliable reference is required that delivers the *correct* actual position of the robot. Then, this reference can be used to compare it with the location computed by the new approach, and thus allows assessing the performance of the new method. RouteLoc uses a mixture of a topological and a metric representation. In fact, a typical position estimate would be “the wheelchair is in the segment between junctions J_i and $J_{i'}$ in a distance of, e.g., 256 cm from J_i ”.

A metric self-localization method is used as a reference. To be able to compare the metric positions determined by the reference locator with the junction/distance pair returned by RouteLoc, the real-world position of each junction is determined in advance. Thus, it is possible to compute an (x, y, θ) triple

Table 1. Comparison between RouteLoc and two other localization approaches

Aspect	RouteLoc	Simmons & Koenig [24]	Thrun <i>et al.</i> [27]
sensor input	odometry (+ 2 sonars for generalization)	odometry + sonars	odometry + camera or laser range finder
setting	campus (in-/outdoor)	indoor office environment	indoor museum
complexity	144 junctions for 46 nodes and 100 edges, depends on number of decision points	3348 Markov states for 95 nodes and 180 edges, depends on extent of environment	About 1000 samples for an indoor environment, number of samples adaptable
memory	very low	very low	huge
precision	Position estimate given by junction and metric offset in the corresponding corridor	Topological map is represented by a set of Markov states (resolution 1m, 90° orientation steps)	Samples indicate position, only small errors

from the junction/distance representation that can be compared to the metric position returned by the reference locator.

6.1 Scan Matching

The method used as a reference was developed by Röfer [19] and is based on earlier work by Kollmann and Röfer [10]. They improved the method of Weißet al. [29] to build maps from measurements of laser range sensors (laser scanners) using a histogram-based correlation technique to relate the individual scans. They introduced state-of-the-art techniques to the original approach, namely the use of projection filters [13], line-segmentation, and multi-resolution matching. The line-segmentation was implemented employing the same approach that was already used for route generalization presented in Sect. 2.1. It runs in linear time with respect to the number of scan points and is therefore faster than other approaches, e. g. the one used by Gutmann and Nebel [8].

The generation of maps is performed in real-time while the robot moves. An important problem in real-time mapping is consistency [13], because even mapping by scan-matching accumulates metric errors. They become visible when a loop is closed. Röfer [19,20] presented an approach to self-localize and to map in real-time while keeping the generated map consistent.

6.2 Experimental Setup

Experiments with the Bremen Autonomous Wheelchair “Rolland” have been carried out on the campus of the Universität Bremen (cf. Fig. 12a) . The wheelchair was driven indoors and outdoors along the dashed line shown in Fig. 12a, visited seven different buildings and passes the boulevard which connects the buildings.

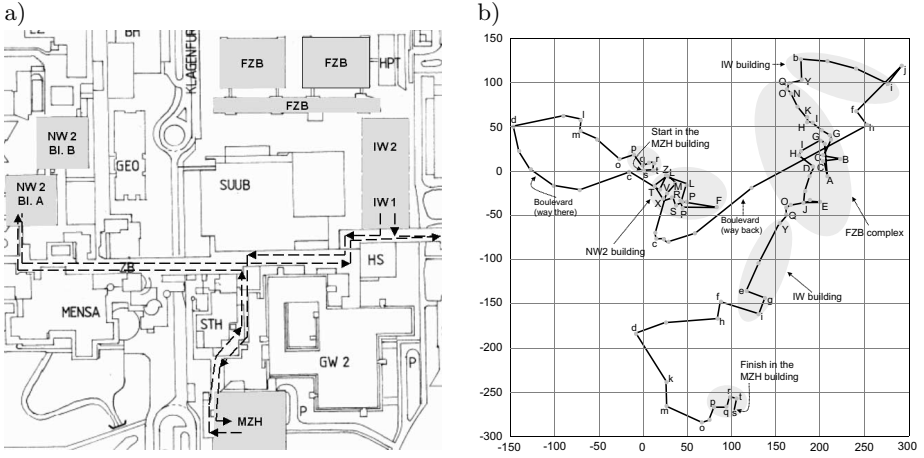


Fig. 12. a) The campus of the Universität Bremen (380m $\times$ 322m). b) Route generalization of odometry data recorded on the campus.

The traveled distance amounts to 2,176m. Traveling along this route with a maximum speed of 84cm/s takes about 75min. While traveling, the wheelchair generated a log file which recorded one state vector every 32ms. Such a state vector contains all the information available for the wheelchair: current speed and steering angle, joystick position, current sonar measurements, and complete laser scans. As mentioned, only locomotion data and the measurements of two sonar sensors are used for the self-localization approach presented here. Feeding the log file (192MB) into the simulator SimRobot [17], it is possible to test the algorithm with *real* data in a simulated world. Note that the simulator works in real-time, i. e. it also delivers the recorded data in 32ms intervals to the connected software modules, one of which is the self-localization module.

For the evaluation of the approach, a laser-scan map of the whole route was generated, using the scan matching method presented in [19]. For such a large scene, the laser map deviates from the original layout of the environment in that the relative locations of the buildings are not 100% correct. Therefore, the route-graph was embedded into the laser scan map making it possible to compare both localization results on a metric basis while traveling through the route with simultaneously active scan matching and route localization modules.¹ It consists of 46 graph nodes and 144 junctions. The represented corridors range in length from 4.3m to 179m.

The deviations between the metric positions determined by the reference locator and the locations calculated by RouteLoc are depicted in Fig. 14. Note that the horizontal axis corresponds to the travel *time* along the route and not to travel *distance*, i. e. the wheelchair stopped several times and also had to shunt

¹ That is the reason why the layout of the route graph depicted in Fig. 13b differs from the map shown in Fig. 12a.

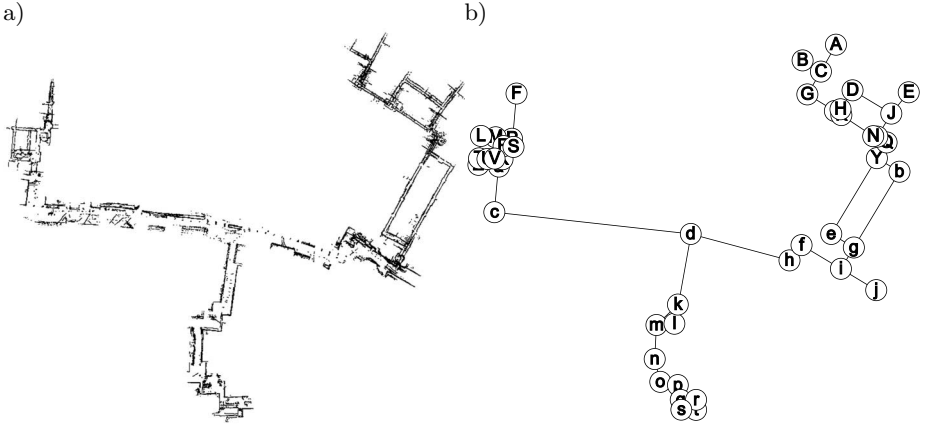


Fig. 13. a) Laser map generated along the route depicted in Fig. 12a. b) Route graph representing the relevant part of the campus.

sometimes, so that distances along this axis do not directly correspond to metric distances along the route.

As **RouteLoc** represents the environment as edges of a graph, its metric precision is limited. The edges of the route graph are not always centered in the corridors; therefore, deviations perpendicular to a corridor can reach its width, which can be more than 10 m outdoors (e. g. corridor *dc*). There are three reasons for deviations along a corridor: first, they can result from the location at which the current corridor was entered (see Sect. 3.3). The bandwidth of possibilities depends on the width of the previous corridor. Second, deviations can be due to odometry errors, because the wheelchair can only correct its position when it drives around a corner. In case of the boulevard (corridor *cdh*), the wheelchair has covered approximately 300 m without the chance of re-localization. Third, deviations can also result from a certain delay before a turn is detected (e. g. the peak after *JE* in Fig. 14). Such generalization delays are discussed in Sect. 3.3 and are also the reason for some peaks such as the one at the end of the boulevard (*dc*).

Even though the odometry data turned out to be very bad (see Fig. 12b), the approach presented here is able to robustly localize the wheelchair. It takes a while before the initial uniform distribution adapts in such a way that there is sufficient confidence to pose a reliable hypothesis about the current position of the robot. But if this confidence is once established, the position is correctly tracked.

7 Conclusion and Future Work

Self-Localization of mobile robots in large-scale environments can be efficiently realized if a hybrid representation of the environment is used. The probabilistic

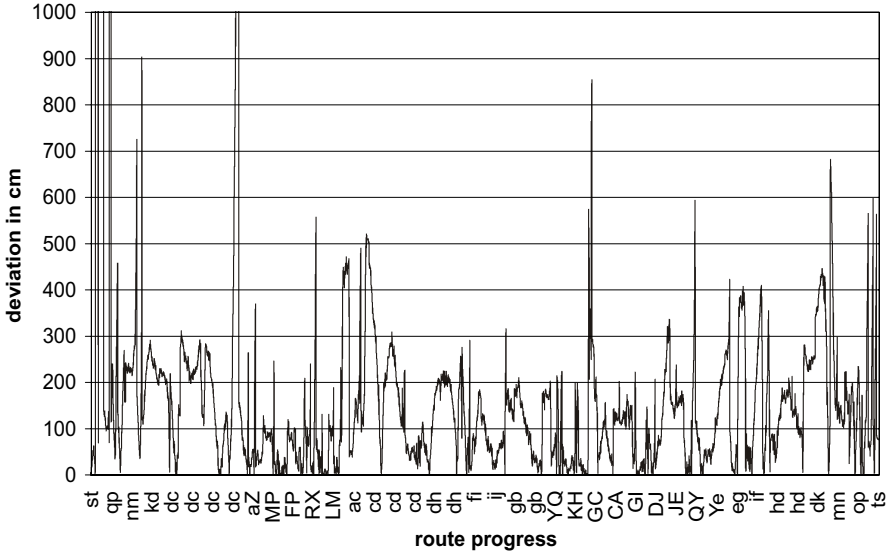


Fig. 14. Deviations of RouteLoc’s position estimates from those made by the laser scan based localization. The letters correspond to segments between the junction labels used in Fig. 13b, but due to the lack of space, some are missing.

approach presented here matches an incremental generalization of the traveled route with an integrated topological-metric map, the *route graph*. Real-world experiments at the Universität Bremen showed the robustness and efficiency of the algorithm. RouteLoc needs only very little input (only odometry data). It is fast and well-scaling, but is sometimes not as precise as other (metric) approaches. Therefore, it should be regarded as a basic method for absolute self-localization that can be extended on demand. In the first place, a disambiguation of situations and the resulting reduced time for the initial localization can be obtained if the route generalization and the route graph were augmented by feature vectors. Additional sensors to detect the features as well as dialogs with the human driver will help here.

RouteLoc will be extended such that self-localizing becomes possible even in a-priori unknown environments (SLAM). For this purpose, the robot has to build the route graph from scratch during runtime and, subsequently, it has to solve the problem of place integration. That means, it has to find out whether its current position is already represented in the route graph, or whether it is located in a corridor that is so far unknown.

Acknowledgements

The Deutsche Forschungsgemeinschaft supports this work through the priority program “Spatial Cognition”.

References

1. J. Borenstein, H. R. Everett, and L. Feng. *Navigating Mobile Robots – Systems and Techniques*. A.K. Peters, Ltd., USA, 1996.
2. W. Burgard, D. Fox, and D. Henning. Fast grid-based position tracking for mobile robots. In G. Brewka, Ch. Habel, and B. Nebel, editors, *KI-97: Advances in Artificial Intelligence*, Lecture Notes in Artificial Intelligence, pages 289–300, Berlin, Heidelberg, New York, 1997. Springer.
3. H. Choset and K. Nagatani. Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization. *IEEE Transactions on Robotics and Automation*, 17(2):125 – 136, April 2001.
4. A. Elfes. Occupancy grids: A stochastic spatial representation for active robot perception. In S. S. Iyengar and A. Elfes, editors, *Autonomous Mobile Robots*, volume 1, pages 60–70, Los Alamitos, California, 1991. IEEE Computer Society Press.
5. S. P. Engelson and D. V. McDermott. Error correction in mobile robot map learning. In *Proceedings of the IEEE Int.'l Conf. on Robotics and Automation*, pages 2555–2560, Nice, France, May 1992. IEEE.
6. C. Eschenbach, C. Habel, L. Kulik, and A. Leßmöllmann. *Shape Nouns and Shape Concepts: A Geometry for 'Corner'*, volume 1404 of *Lecture Notes in Artificial Intelligence*, pages 177–201. Springer, Berlin, Heidelberg, New York, 1998.
7. D. Fox, W. Burgard, F. Dellaert, and S. Thrun. Monte Carlo localization: Efficient position estimation for mobile robots. In *Proc. of the National Conference on Artificial Intelligence*, 1999.
8. J.-S. Gutmann and B. Nebel. Navigation mobiler Roboter mit Laserscans. In P. Levi, Th. Bräunl, and N. Oswald, editors, *Autonome Mobile Systeme*, Informatik aktuell, pages 36–47, Berlin, Heidelberg New York, 1997. Springer.
9. J.-S. Gutmann, T. Weigel, and B. Nebel. A fast, accurate, and robust method for self-localization in polygonal environments using laser-range-finders. *Advanced Robotics*, 14(8):651 – 668, 2001.
10. J. Kollmann and T. Röfer. Echtzeitkartenaufbau mit einem 180°-Laser-Entfernungssensor. In R. Dillmann, H. Wörn, and M. von Ehr, editors, *Autonome Mobile Systeme 2000*, Informatik aktuell, pages 121–128. Springer, 2000.
11. B. Kuipers, R. Froom, Y. W. Lee, and D. Pierce. The semantic hierarchy in robot learning. In J. Connell and S. Mahadevan, editors, *Robot Learning*, pages 141–170. Kluwer Academic Publishers, 1993.
12. A. Lankenau and T. Röfer. The Bremen Autonomous Wheelchair – a versatile and safe mobility assistant. *IEEE Robotics and Automation Magazine*, “Reinventing the Wheelchair”, 7(1):29 – 37, March 2001.
13. F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4:333–349, 1997.
14. A. Mojaev and A. Zell. Online-Positionskorrektur für mobile Roboter durch Korrelation lokaler Gitterkarten. In H. Wörn, R. Dillmann, and D. Henrich, editors, *Autonome Mobile Systeme*, Informatik aktuell, pages 93–99, Berlin, Heidelberg, New York, 1998. Springer.
15. A. Musto, K. Stein, A. Eisenkolb, and T. Röfer. Qualitative and quantitative representations of locomotion and their application in robot navigation. In *Proc. of the 16th International Joint Conference on Artificial Intelligence (IJCAI-99)*, pages 1067–1073, San Francisco, CA, 1999. Morgan Kaufman Publishers, Inc.

16. I. Nourbakhsh, R. Powers, and S. Birchfield. Dervish: An office-navigating robot. *AI Magazine*, 16:53–60, 1995.
17. T. Röfer. Strategies for using a simulation in the development of the Bremen Autonomous Wheelchair. In R. Zobel and D. Moeller, editors, *Simulation-Past, Present and Future*, pages 460–464. Society for Computer Simulation International, 1998.
18. T. Röfer. Route navigation using motion analysis. In *Proc. Conf. on Spatial Information Theory '99*, volume 1661 of *Lecture Notes in Artificial Intelligence*, pages 21–36, Berlin, Heidelberg, New York, 1999. Springer.
19. T. Röfer. Building consistent laser scan maps. In *Proc. of the 4th European Workshop on Advanced Mobile Robots (Eurobot 2001)*, volume 86 of *Lund University Cognitive Studies*, pages 83 – 90, 2001.
20. T. Röfer. Konsistente Karten aus Laser Scans. In *Autonome Mobile Systeme 2001*, Informatik aktuell, pages 171–177. Springer, 2001.
21. T. Röfer and A. Lankenau. Ensuring safe obstacle avoidance in a shared-control system. In J. M. Fuertes, editor, *Proc. of the 7th Int. Conf. on Emergent Technologies and Factory Automation*, pages 1405 – 1414, 1999.
22. T. Röfer and A. Lankenau. Architecture and applications of the Bremen Autonomous Wheelchair. *Information Sciences*, 126(1-4):1 – 20, July 2000.
23. J.S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, New Jersey, USA, 1995.
24. R. Simmons and S. Koenig. Probabilistic robot navigation in partially observable environments. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI-95*, pages 1080–1087, 1995.
25. S. Thrun. Learning maps for indoor mobile robot navigation. *Artificial Intelligence*, 99:21 – 71, 1998.
26. S. Thrun, W. Burgard, and D. Fox. A Real-Time Algorithm for Mobile Robot Mapping With Applications to Multi-Robot and 3D Mapping. In *Proc. of the IEEE Int. Conf. on Robotics & Automation*, pages 321 – 328, 2000.
27. S. Thrun, D. Fox, W. Burgard, and F. Dellaert. Robust Monte Carlo localization for mobile robots. *Artificial Intelligence*, 101:99 – 141, 2000.
28. N. Tomatis, I. Nourbakhsh, and R. Siegwart. Simultaneous localization and map building: A global topological model with local metric maps. In *Proceedings of the IEEE/RSJ Int.'l Conf. on Intelligent Robots and Systems (IROS 2001)*, Maui, Hawaii, October 2001.
29. G. Weiß, C. Wetzler, and E. von Puttkamer. Keeping Track of Position and Orientation of Moving Indoor Systems by Correlation of Range-Finder Scans. In *Proc. Int. Conf. on Intelligent Robots and Systems 1994 (IROS-94)*, pages 595–601, 1994.
30. S. Werner, B. Krieg-Brückner, and Th. Herrmann. *Modelling Navigational Knowledge by Route Graphs*, volume 1849 of *Lecture Notes in Artificial Intelligence*, pages 295–316. Springer, Berlin, Heidelberg, New York, 2000.
31. D. van Zwynsvoorde, T. Simeon, and R. Alami. Incremental topological modeling using local Voronoï-like graphs. In *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and System (IROS 2000)*, volume 2, pages 897 – 902, Takamatsu, Japan, October 2000.
32. D. van Zwynsvoorde, T. Simeon, and R. Alami. Building topological models for navigation in large scale environments. In *Proc. of IEEE Int. Conf. on Robotics and Automation ICRA 2001*, pages 4256 – 4261, Seoul, Korea, May 2001.

The Role of Geographical Slant in Virtual Environment Navigation

Sibylle D. Steck¹, Horst F. Mochnatzki², and Hanspeter A. Mallot²

¹ DaimlerChrysler Research & Technology, Ulm, Germany

² Dept of Zoology, University of Tübingen, Germany.

Abstract. We investigated the role of geographical slant in simple navigation and spatial memory tasks, using an outdoor virtual environment. The whole environment could be slanted by an angle of 4° . Subjects could interact with the virtual environment by pedaling with force-feedback on a bicycle simulator (translation) or by hitting buttons (discrete rotations in 60° steps). After memory acquisition, spatial knowledge was accessed by three tasks: (i) pointing from various positions to the learned goals; (ii) choosing the more elevated of two landmarks from memory; (iii) drawing a sketch map of the environment. The number of navigation errors (wrong motion decisions with respect to the goal) was significantly reduced in the slanted conditions. Furthermore, we found that subjects were able to point to currently invisible targets in virtual environments. Adding a geographical slant improves this performance. We conclude that geographical slant plays a role either in the construction of a spatial memory, or in its readout, or in both.

1 Introduction

When we find our way in a familiar environment, we use various cues or types of information to find out where we are and, more importantly, where we should head from there. Besides egomotion information, which can be used for path integration, objects and landscape configurations are the most important sources of information. Places can be characterized by recognized objects (local landmarks) or by geometrical peculiarities such as the angle under which two streets meet (cf. Gouteux & Spelke 2001). A mixture of place and geocentric direction information is provided by distant or global landmarks (cf. Steck and Mallot 2000 for a discussion of local and global landmarks). Finally, true geocentric direction (or compass) information is conveyed by cues like the azimuth of the sun (in connection with the time of day) or the slant direction of a ramp-like terrain.

So far, the role of geographical slant and elevation in navigation is only poorly understood. Creem and Proffitt (1998) asked subjects to adjust the slant of a board to previously seen slants of terrain and found that slants as low as 4 degrees are well perceived. In an earlier study, Proffitt et al. (1995) showed that in virtual environments, subjects are also able to reproduce accurately geographical slant (5° to 60° in 5° steps) on a tilt board. Further, the judgments in the virtual

environments and the naturally presented slants do not differ significantly. This result was confirmed in a study by Proffitt et al. (2001). The memory for elevation of places was studied by Gärling et al. (1990) who showed that subjects were able to judge from memory which of two places in a familiar environment was higher elevated. Subjects who had less experience with the environment tended to exaggerate the elevation differences. Evidence for the use of slant, i.e. the elevation gradient, in human spatial cognition comes from linguistic studies in people living in landscapes with conspicuous slants. Brown and Levinson (1993) and Levinson (1996) report that the Tzeltal language spoken in parts of Mexico uses an uphill/downhill reference frame even in contexts where English or other languages employ a left/right scheme.

In rats, a direct demonstration for the use of slant as a cue to navigation has been provided by Moghaddam et al. (1996). When searching a food source on top of an elevated cone, rats were able to navigate a more direct path than on a flat surface.

Theoretically, there are good reasons to expect that geographical slant should be used in navigation. First, some important navigation tasks such as “find water” can be solved by simply walking downhill. Note that no self-localization is required in this case. Second, geographical slant can provide geocentric¹ compass information which is known to be of great importance in path integration (see Maurer and Séguinot 1995, Mallot 2000). While path integration is principally possible by pure vector summation without any compass, error accumulation is greatly reduced if independent compass information is available. Insects which make extensive use of path integration (Müller and Wehner 1988) obtain compass information from the polarization pattern of the sky light (Rossel 1993). Finally, geographical slant might also act as a local cue characterizing a place. Indeed, it seems quite likely that the same landmark appearing on top of a mountain or halfway along the ascent are readily distinguished. Again, it has been shown in insects, that the so-called snapshot, a view of the environment characteristic of the location is was viewed from, is registered to a compass direction (Cartwright and Collett 1982).

In this paper, we address the question whether global geographical slant can be used by human navigators to improve their performance. Three versions of a virtual environment differing only in the overall slant of the terrain were generated. After exploring one of these environments, subjects’ performance and spatial representation was assessed by measuring the overall navigation performance, the quality of pointing to remembered targets, the quality of judging which of two remembered places was higher in elevation, and the orientation of sketch map drawings.

¹ The term “geocentric” is used to indicate that some information is given in an observer independent coordinate system, fixed to some anchor point in the world. In contrast, the term “geographical” is used only in connection with the word “slant” to indicate that we are talking about the slant of landscapes rather than the slant of object surfaces. Finally, the term “geometrical” refers to depth as local position information, e.g. “a junction where streets meet at an angle of 45 degrees”.

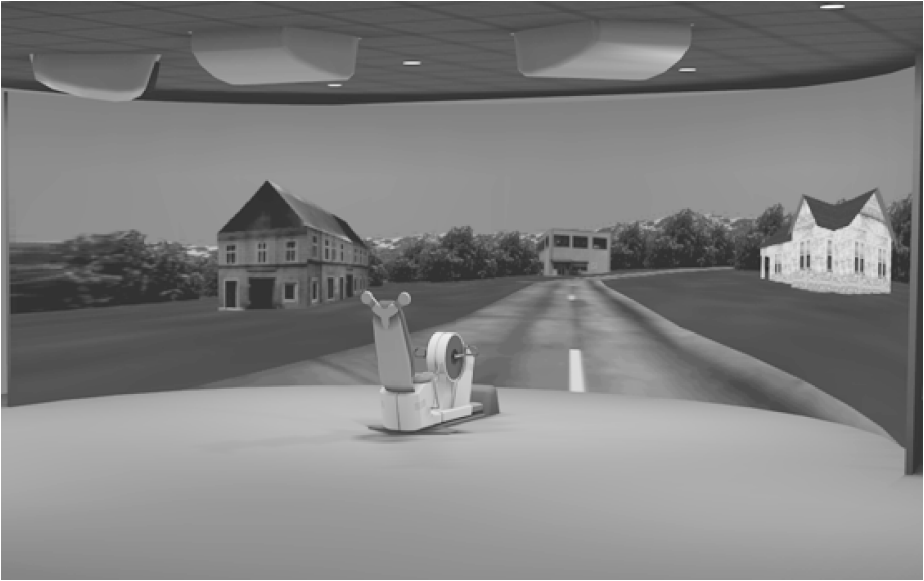


Fig. 1. Virtual Environments Lab with 180° projection screen showing the Hexatown simulation. The subject was seated on a virtual reality bicycle in the center of the half cylinder.

2 Method

2.1 Subjects

A total of 36 subjects (18 male and 18 female, aged 15–31 years) took part in the experiment. Participation in this experiment was voluntarily and a honorarium was paid for participation.

2.2 Virtual Environment

Graphical Apparatus. The experiment was performed on a high end graphics computer (Silicon Graphics Inc. ONYX2 3-pipe Infinite Reality), running a C-Performer application that we designed and programmed. The simulation was displayed non-stereoscopically, with an update rate of 36 Hz, on a half-cylindrical projection screen (7m diameter and 3.15m height, Fig. 1). The computer rendered three 1280 × 1024 pixel color-images projected side by side with a small overlap. Images were corrected for the curved surface by the projectors to form a 3500 × 1000 pixel display. For an observer seated in the center of the cylinder (eye height 1.25m), this display covered a field of view of 180° horizontally times 50° vertically. The field of view of the observer was identical to the field of view used for the image calculations. A detailed description of the setup can be found in Veen et al. (1998).

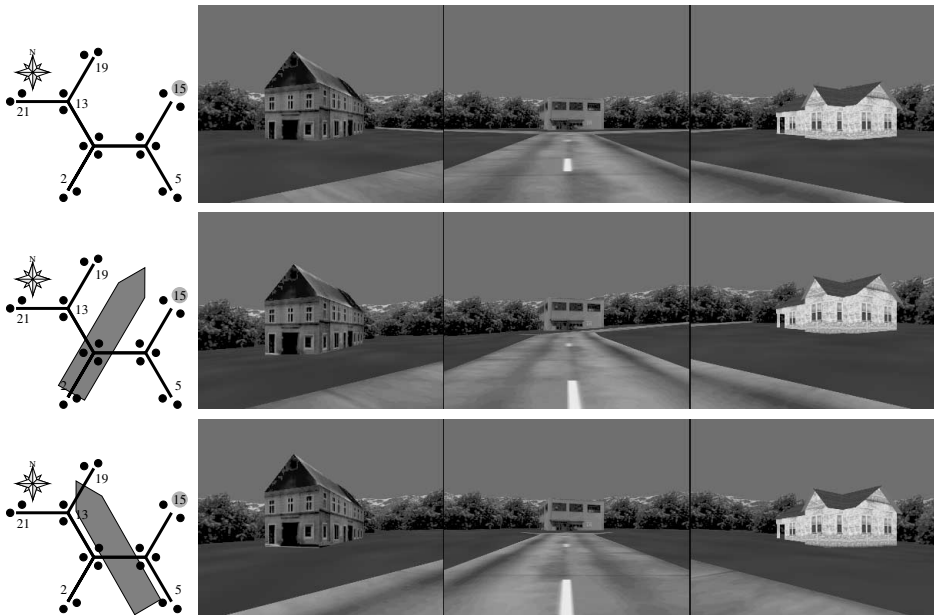


Fig. 2. Overview of the three conditions. **Left:** map of the environments. Landmarks indicated by numbers have been used as goals in the exploration phase and as targets in the pointing phase. **Right:** subjects perspective. Each row shows the three pictures projected on the 180° screen. The images are projected with a small overlap; therefore the discontinuities visible here are not present in the actual experiment. The picture shows the view from the place with object 5 in the direction of the street towards the only adjacent place. **Top row** shows the *Flat* slant condition. **Middle row** shows the *Northeast* slant condition. **Bottom row** shows *Northwest*.

Scenery. In this experiment, we used three similar environments varying only in geographical slant (Fig. 2). In the control condition, the environment was on a flat plane (*Flat*). In the two other conditions, the environment had a global geographical slant with a slant angle of 4° . (The slant angle is the angle between the surface normal and the vertical; a slant angle of 4° is equivalent to an inclination of 7%.) In pilot studies (Mochnatzki 1999), we found that the simulated slant was well above the detection threshold for geographic slant in the same experimental setup. The slanted environments differed in the orientation of the slant with respect to an arbitrarily chosen “North” direction. In one condition, the geographical slant was oriented in the direction of Northeast (*NE*). In a further condition, the slant was to the Northwest (*NW*). The reasons for using two slanted environments are the following: First, the street raster of the virtual town is not completely isotropic, so different slant directions might have different effects. Second, the sequence of learning tasks used in the exploration phase (see below) introduces an additional anisotropy which makes it necessary to use at least two slant directions.

The model of the environment was generated using MultiGen 3-D modeling software. The environment consisted of an octagonal ground plane surrounded by a flat background showing a regular mountain range. The buildings were constructed using Medit 3-D modeling software. Schematic maps of the town appear in the left side column of Fig. 2. Maps or aerial views were not shown to the subjects. The virtual environment (called “Hexatown”, see Gillner and Mallot, 1998, Steck and Mallot 2000, and Mallot and Gillner 2000) consisted of a hexagonal raster of streets with a distance of 100 meters between adjacent junctions. A junction was built of three adjoining streets forming 120° corners. In each corner, an object (building, gas station, etc.) was placed, see Fig. 2. At the periphery of Hexatown, streets ended blindly. These dead-ends were marked by barriers 50 meters from the junction. A circular hedge or row of trees was placed around each junction with an opening for each of the three streets (or dead ends) connected to that junction. This hedge looked the same for all junctions and prevented subjects from seeing the objects at distant junctions.

The usage of geometrical cues, as demonstrated, e.g., by Hermer and Spelke (1994) and Janzen et al. (2000) is not possible in Hexatown. All junctions are identical and symmetrical, so that when approaching a junction, one cannot infer the approach direction nor the approached place from the geometrical layout. As compared to rectangular city rasters, which are also symmetrical, the hexagonal layout has the advantage that there is no straight-on direction that might be preferred over the branching streets.

Interaction. Subjects navigated through Hexatown using a virtual reality bicycle (a modified versions of a training bicycle from *CyberGearTM*) which can be seen in Fig. 1 (for details see van Veen et al. 1998). The bicycle has force-feedback, i.e. when pedaling uphill, the subjects have to exert more force than when cycling downhill. The setup thus provides both visual and proprioceptive slant information.

At the junctions, 60° turns could be performed by pressing one of two buttons (left or right) fixed to the bicycle. Once the simulated turn movement was initiated, it followed a predefined velocity profile: turns took 4 seconds, with a maximum speed of 30° per second and symmetric acceleration and deceleration (ballistic movement). The smooth profiles for rotation were chosen to minimize simulator sickness. Translations on the street were initiated by pressing an additional button. Translations were not ballistic; translation velocity was controlled by the pedal revolution, using a mechanical motion model that took into account the current geographical slant. Subjects could only influence the speed, but were not able to change the direction, i.e. they were restricted to the streets.

In the sequel, the motion initiated by pressing the buttons will be referred to as motion decision.

2.3 Procedure

The three experimental conditions were tested in an between-subject design, using 12 subjects per condition. Subjects were run through the experiment individually.

The experiment had four different phases: a navigation phase, pointing judgments, elevation comparison, and map drawing. In the navigation task, the subjects had to find a previously shown goal using the shortest possible path. The navigation phase consisted of 15 search tasks. In the pointing judgment, subjects were asked to carry out directional judgments to previously learned goals. In the elevation judgments, subjects had to choose which learned goal was higher up in the environment. This part was omitted in the *Flat* condition. Finally, subjects had to draw a map from the learned environment. For each part, subjects were instructed separately. Therefore, they were uninformed of all tasks in advance. On average, subjects needed 90 min for all tasks.

Navigation Phase. First, the subjects had to solve 15 search tasks in the primarily unknown environment (Fig 2). Before each trial, a full 180° panoramic-view at the goal location was shown. By pressing a button on the handles of the VR-bicycle, the goal presentation was terminated and subjects were positioned at the current starting position. When they had reached their goal, a message was displayed, indicating whether they had used the path with the least number of motion decisions (“fastest” path), or not. The task was repeated until it was first completed without mistakes. During the entire navigation phase, the subjects had the possibility to expose a small picture of the current goal object on a gray background in the bottom left corner of the middle screen by pressing a special button. The starting point of the first five tasks was landmark 15 (*home*). The solutions of the first five search tasks covered the entire maze; we therefore call these phase *exploration*. The next ten routes were either *return paths* from the previously learned goals to the landmark 15, or *novel paths* between the goals, which were learned in the exploration phase. Search tasks involving a return and a novel path were carried out in alternation. The navigation phase ensured that all subjects reached a common fixed performance level for the subsequent pointing judgments.

Pointing Judgments. Pointing judgments were made to evaluate the internal representation of the learned environment. The subjects were placed in front of a learned goal, which was randomly chosen. They were asked to orient themselves towards one of four other goals (except *home*) by continuously turning the simulated environment. A fixed pointer (fixed with respect to the screen) was superimposed on the turning image to mark the forward direction to which the goal had to be aligned. Note that this procedure differs considerably from pointing in real environments or in virtual environments presented using a head-mounted display, in that the observer’s arm or body need not move. All that moves during the “pointing” judgment is the image of the simulated environment. For a discussion of pointing procedures, see Montello et al. (1999). Altogether, the subjects had to point to twenty goals. One of these goals was directly visible from one of the reference points. This pointing task was therefore excluded from further analysis.

Elevation Judgments. In order to test whether elevation information was also stored, elevation judgments were collected in the *Northeast* and *Northwest* conditions. Pictures of two goals of different elevation were presented in isolation

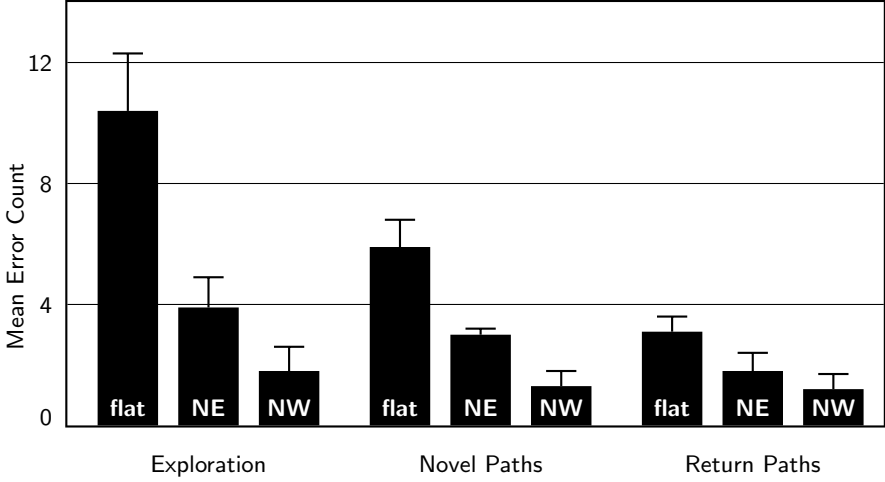


Fig. 3. Mean error count in the navigation phase. Mean number of errors for the three route types (*exploration*, *novel paths*, and *return paths*) and the three slant conditions, *flat*, *slanted NE*, and *slanted NW*. The error bars present one standard error of the mean.

on a gray screen and the subjects had to decide as accurately and as quickly as possible, which goal had appeared at higher elevation in the training environment. For each of the two slant conditions, ten pairs of goals were selected and tested.

Map Drawing. In the final phase of the experiment, subjects were asked to draw by hand as detailed a map of the test environment as possible. They were given a pen and a paper. The paper had a printed frame to restrict their drawings. There was no time limit for the subjects.

3 Results

3.1 Errors in the Navigation Phase

In the navigation phase, the trajectories of the subjects for every search task were recorded. Every movement decision that did not reduce the distance to the goal was counted as an error. Figure 3 shows the mean number of errors per path type (*exploration*, *return paths*, and *novel paths*) and per slant condition. A three-way ANOVA (3 path types $\times$ 3 slant conditions $\times$ gender) shows a significant main effect of slant condition ($F(2, 30) = 5.78$, $p = 0.008^{**}$). As figure 3 shows, more errors were made in the *Flat* condition than in the *Northeast* condition. In the *Northwest* slant condition, the least amount of error was made. Further, there was a highly significant main effect of the path type ($F(2, 60) = 27.69$, $p < 0.001^{***}$). In all three slant conditions, the largest number of errors occurred in the *exploration* phase (first five paths, all starting from home). The second

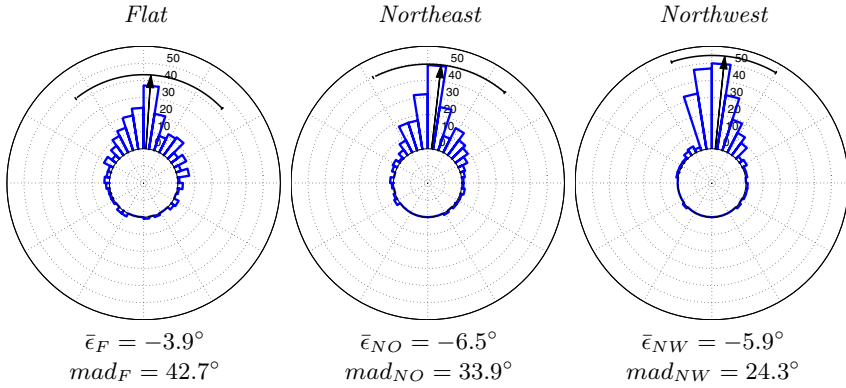


Fig. 4. Pointing Error. Circular plots for the slant condition Flat, Northeast and Northwest. $\bar{e}$: circular mean of the error (arrow). mad : mean angular deviation (segment).

largest number of errors was made for the *novel paths* (connection paths between goals, none of which was the home), while the *return paths* were navigated with the smallest number of errors. Note that the *return paths* alternated with the *novel paths* in the task sequence; therefore the difference in the number of errors of these two path types cannot be explained by differences in the time spent in the environment before each task.

A significant interaction between slant condition and path type was also found ($F(4, 60) = 4.37$, $p = 0.004^{**}$). It may reflect a floor effect for the *Northwest* slant condition. Since the number of errors was very small in this condition anyway, the effects of condition and path type do not completely superimpose. No difference in the mean number of errors was found between male and female subjects (men: 11.5 ± 1.9 , women: 10.2 ± 1.6 , $F(1, 30) = 0.300$, $p = 0.59$ n.s.).

3.2 Pointing Judgments

The pointing judgments were stored as angles in degrees with respect to the arbitrarily chosen North direction. Since pointing judgments are periodic data (e.g., 181° is the same direction as -179°), we used circular statistics (see Batschelet 1981) to analyze pointing judgments. The circular means ($\bar{a}$) were calculated by summing the unit vectors in the direction of the pointings. The resultant vector was divided by the number of averaged vectors. The length of the mean vector is a measure for the variability of the data.

To compare the different slant conditions, the deviations from the correct values were averaged over all tasks. Figure 4 shows the deviation from the correct values for all tasks and all subjects. The measured values were distributed in 9° bins. The arrow shows the direction of the circular mean of the errors. The length of the mean vectors is inversely proportional to the “mean angular deviation” shown as circular arc in Fig. 4. The mean vectors are close to zero for all conditions, as is to be expected since we plotted the pointing error.

Table 1. Comparison of the variances of pointing in the three slant conditions.

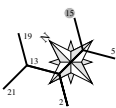
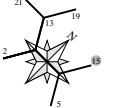
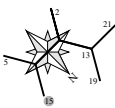
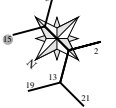
Comparison	$F(198,198) = \frac{mad_1^2}{mad_2^2}$	p
Flat– Northeast	1.72	< 0.001 ***
Flat – Northwest	3.5257	< 0.001 ***
Northeast – Northwest	2.0408	< 0.001 ***

For comparing the variances of the different slant conditions, we compared the arithmetic mean of the squares of the mean angular deviation of each subject using the circular F-test (Batschelet 1981, chap.6.9). There is a highly significant difference between all conditions, see Table 1.

3.3 Elevation Judgment

In this part, subjects in the *slanted NE* and *slanted NW* conditions were tested to determine, if they stored the relative elevations of the objects. The subjects in the *Northwest* slant condition gave 109 correct answers out of 120, 90.8%, and the subjects in *Northeast* 94 correct answers out of 120, 78.3%. The answers of the subjects differed significantly from a binomial distribution with $p = 50\%$ which would imply pure guessing ($\chi^2_{NE}(10) = 492.0, p < 0.001^{***}, \chi^2_{NW}(10) = 3838.9, p < 0.001^{***}$). Therefore, we conclude that the subjects were able to differentiate object elevation. The percentage correct of the *Northwest* condition was significantly higher than the percentage *Northeast* (U-Test after Mann and Whitney $U(12, 12, p = 0.05) = 37, p \leq 0.05^*$).

Table 2. Alignment of sketch maps in the three slant conditions

	Alignment				
	NE-up	NW-up	SW-up	SE-up	ambiguous
					
<i>flat</i>	3	0	6	0	3
<i>slanted NE</i>	6	0	2	0	4
<i>slanted NW</i>	2	5	1	0	4

3.4 Map Drawings

The map drawings were used to study how subjects implemented the geographical slant in their representation. Single maps were mostly quite good, since the

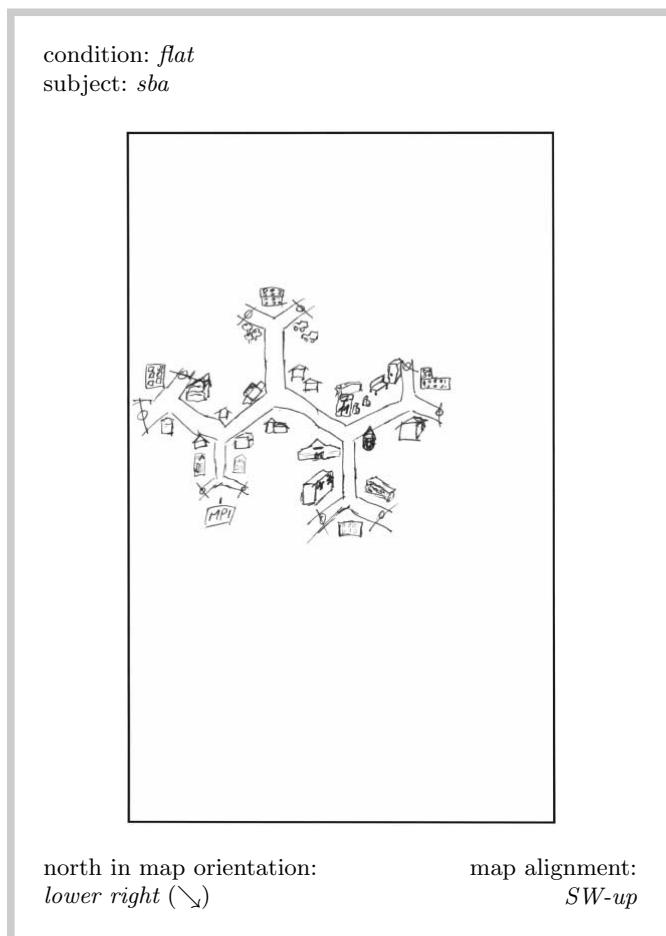
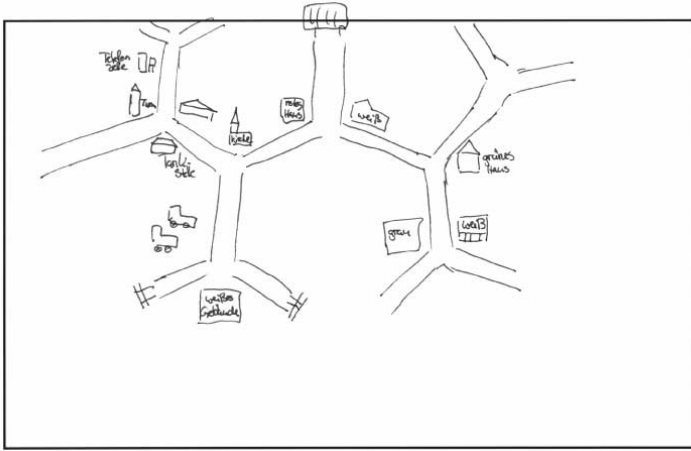


Fig. 5. Sketch map drawn by subjects *sba* in condition *flat*. The drawing is aligned in the sense that all buildings are given in perspective with the same vantage point. The top of the page corresponds to Southwest. The bold gray box indicates the size of the sketching paper ($A4 = 21\text{ cm} \times 29.7\text{ cm}$). The thin black box is the frame printed on the sketching paper to prevent subjects from starting their drawings too closely to the edge of the paper.

geometry of the junction was often correctly depicted. Only three out of thirty-six subjects drew all junctions as right angle junctions. Four further subjects drew right angles at some junctions. All except one very sparse map, contained object 15, which was the start point of the first five routes.

We were interested in whether the slant conditions influenced the map drawings. Therefore, all maps were examined for alignment. A map was considered “aligned”, if either a uniform orientation of lettering (e.g., Fig. 7) or a perspective of the drawn objects (e.g., Fig. 5) was apparent to the authors. Judgments

condition: *slanted NE*
 subject: *spe*



north in map orientation:
 upper left (↖)

map alignment:
 NE-up

Fig. 6. Sketch map drawn by subject *spe* in condition *Northeast*. The alignment is apparent from the drawings of the houses. The top of the page corresponds to Northeast, i.e. the more elevated locations are drawn more towards the top of the page. The boxes represent the margin and inner frame of the sketching paper (cf. Fig. 5).

of alignment were carried out independently and maps judged differently are labeled “ambiguous” in Table 2.

The maps were categorized in four groups: NE-up, SE-up, SW-up and NW-up. Table 2 lists the number of drawn maps for all alignment categories for the three different slant conditions (*flat*, *slanted NE*, and *slanted NW*). In the *flat* slant condition, the SW-up alignment was found six times. In this alignment category, object 15 is at the lower edge of the map, and the street, which leads to the next junction, points to the top (cf. Figure 5). Further, the category NE-up (in which the object 15 is at the top edge of the map, and the street, which leads to the next junction, points to the bottom) occurred three times. In the *Northeast* slant condition, the alignment category NE-up occurred six times and SW-up two times. In both cases (NE-up, SW-up), the maps were aligned with the gradient along the geographical slant, with the majority of the maps aligned to the uphill gradient (see Figure 6). In the *Northwest* slant condition, the alignment category NW-up (i.e., uphill along the gradient) occurred five times (cf. Figure 7). There were two maps of the category NE-up and one map of the category SW-up. The distributions of the maps in the alignment categories differ significantly ($\chi^2(\text{slanted NW/flat}) = 30.5$,

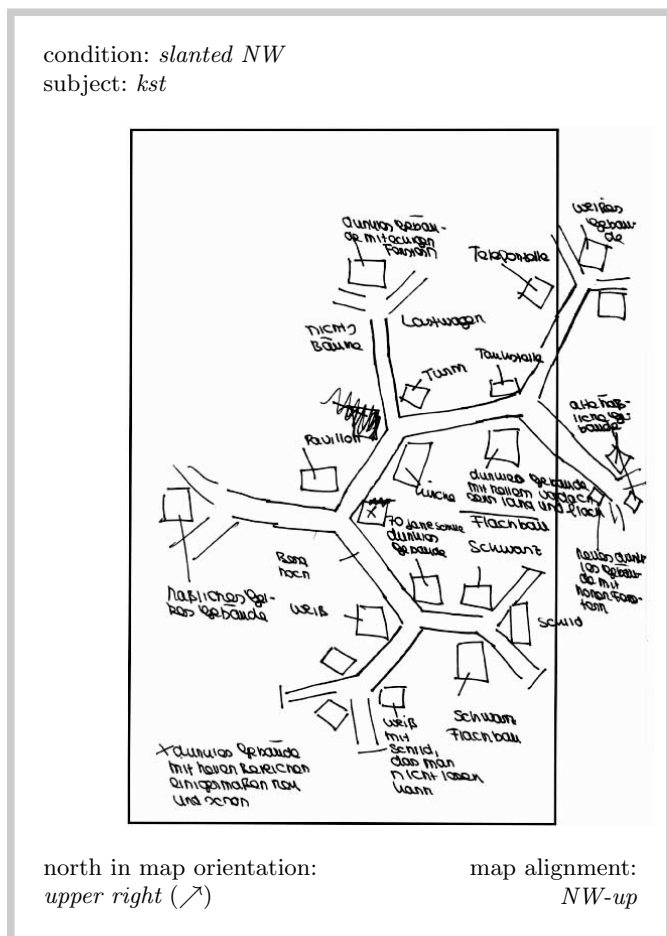


Fig. 7. Sketch map drawn by subject *kst* in condition *Northwest*. The alignment is apparent from the lettering (in German). The top of the page corresponds to Northwest, i.e. the more elevated locations are drawn more towards the top of the page. The boxes represent the margin and inner frame of the sketching paper (cf. Fig. 5).

$df = 3, p < 0.001^{***}, \chi^2(\text{slanted NW/slanted NE}) = 14.0, df = 3, p = 0.003^{**}, \chi^2(\text{slanted NE/flat}) = 9.5, df = 3, p = 0.02^{*})$.

4 Discussion

4.1 Navigation Is Improved by Geographical Slant Information

The number of navigation errors in the navigation phase was strongly reduced in the slanted environments (Fig. 3). This result clearly indicates that slant information is used by the subjects. It is important to note that this improvement

occurred for each route type (exploration, return, novel route) individually, not just for routes leading to a goal uphill in the environment. It appears therefore that slant information is used to improve spatial knowledge in general. In contrast, in the study by Moghaddam et al. (1996), only the navigation to targets on top of a hill was addressed.

A surprising result is the difference between the two slant conditions, which differ only in the direction of the slant relative to the maze layout. We speculate that the difference is related to the fact that in the *slanted NE* condition, the longest route (four segments) is running in zigzag pattern up and down the slope, whereas in the *slanted NW* condition, the longest route is constantly going uphill or downhill. Therefore, the slant information is ambiguous in the *slanted NE* condition.

The results from the navigation part of the experiment are well in line with the pointing judgments. Again, pointing is better for the slanted conditions, and it is also better for the *slanted NW* than for the *slanted NE* condition. We found no difference in judgment accuracy between pointings parallel to the slant and pointings perpendicular to the slant.

Improved pointing in slanted environments is to be expected if slant is used as a compass in a path integration scheme. However, this mechanism does not explain the difference found between the two slant conditions.

4.2 Slant Information and Spatial Memory

The results from the elevation judgment part of the experiment show that the subjects remember the relative elevation of the various objects in the maze. This finding is well in line with the results of Gärling et al. (1990). In graph-like, or topological models of spatial memory (Kuipers 2000, Gillner & Mallot 1998), elevation may be attached as a label to each node of the graph. Alternatively, local slant information could be attached to the edges of the graph. In a recent model by Hübner & Mallot (2002), the graph contains local metric information, including distances between adjacent nodes and angles between adjacent edges. A generalized multi-dimensional scaling algorithm is then used to estimate the metric 2D coordinates of each node. This scheme can be generalized to account for slant data. Local slant data could then be used to generate elevation estimates per node, by some sort of 3D path integration. Evidence for 3D path integration in insects has recently been provided by Wohlgenuth et al. (2001).

Overall slant direction, as seems to be present in the map drawings, is not easily represented in pure graph models. Indeed, some metric structure (as opposed to mere neighborhood information) is necessary to represent a global slant direction in a cognitive map. Further experiments with irregular slants will be needed to access the roles of global and local slant directions.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft, Grant Numbers MA 1038/6-1, MA 1038/7-1 and by the Office of Naval Research Grant

N00014-95-1-0573 award to Jack Loomis. We are grateful to Silicon Graphics Inc., Prof. F. Leberl (Univ. Graz), and the Salford University, UK, for providing the VR-models used in this experiments. The authors thank Jan Restat and Pavel Zahorik for comments and suggestions on an earlier draft of this manuscript. We are grateful to Scott Yu for providing the 3D model of our virtual environments lab shown in Fig. 1.

References

- Batschelet, E. (1981). *Circular Statistics in Biology*. Academic Press, London.
- Brown, P. and Levinson, S. C. (1993). "Uphill" and "Downhill" in Tzeltal. *Journal of Linguistic Anthropology*, 3(1):46 – 74.
- Cartwright, B. A. and Collett, T. S. (1982). How honey bees use landmarks to guide their return to a food source. *Nature*, 295:560 – 564.
- Creem, S. H. and Proffitt, D. R. (1998). Two memories for geographical slant: Separation and interdependence of action and awareness. *Psychonomic Bulletin & Review*, 5:22 – 36.
- Gärling, T., Böök, A., Lindberg, E., and Arce, C. (1990). Is elevation encoded in cognitive maps. *Journal of Environmental Psychology*, 10:341 – 351.
- Gillner, S. and Mallot, H. A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience*, 10:445 – 463.
- Gouteux, S. and Spelke, E. S. (2001). Children's use of geometry and landmarks to reorient in an open space. *Cognition*, 81:119 – 148.
- Hermer, L. and Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, 370:57 – 59.
- Hübner, W. and Mallot, H. A. (2002). Integration of metric place relations in a landmark graph. In Dorronsoro, J. R., editor, *International Conference on Artificial Neural Networks (ICANN 2002)*, Lecture Notes in Computer Science. Springer Verlag.
- Janzen, G., Herrmann, T., Katz, S., and Schweizer, K. (2000). Oblique angled intersections and barriers: Navigating through a virtual maze. *Lecture Notes in Computer Science*, 1849:277–294.
- Kuipers, B. (2000). The spatial semantic hierarchy. *Artificial Intelligence*, 119:191 – 233.
- Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic studies. In Bloom, P., Peterson, M. A., Nadel, L., and Garrett, M. F., editors, *Language and Space*, pages 109 – 169. The MIT Press, Cambridge, MA.
- Mallot, H. (2000). *Computational Vision. Information Processing in Perception and Visual Behavior*, chapter Visual Navigation. The MIT Press, Cambridge, MA.
- Mallot, H. A. and Gillner, S. (2000). Route navigation without place recognition: what is recognized in recognition-triggered responses? *Perception*, 29:43 – 55.
- Maurer, R. and Séguinot, V. (1995). What is modelling for? A critical review of the models of path integration. *Journal of theoretical Biology*, 175:457 – 475.

- Mochnatzki, H. (1999). Die Rolle von Hangneigungen beim Aufbau eines Ortsgedächtnisses: Verhaltensversuche in Virtuellen Umgebungen. Diploma thesis, Fakultät für Biologie, Univ. Tübingen.
- Moghaddam, M., Kaminsky, Y. L., Zahalka, A., and Bures, J. (1996). Vestibular navigation directed by the slope of terrain. *Proceedings of the National Academy of Sciences, USA*, 93:3439–3443.
- Montello, D. R., Richardson, A. E., Hegarty, M., and Provenza, M. (1999). A comparison of methods for estimating directions in egocentric space. *Perception*, 28:981 – 1000.
- Müller, M. and Wehner, R. (1988). Path integration in desert ants, *cataglyphis fortis*. *Proceedings of the National Academy of Sciences, USA*, 85:5287 – 5290.
- Proffitt, D. R., Bhalla, M., Gossweiler, R., and Midgett, J. (1995). Perceiving geographical slant. *Psychonomic Bulletin & Review*, 2:409 – 428.
- Proffitt, D. R., Creem, S. H., and Zosh, W. D. (2001). Seeing mountains in mole hills: geographical-slant perception. *Psychological Science*, 12:418 – 423.
- Rossel, S. (1993). Navigation by bees using polarized skylight. *Comparative Biochemistry & Physiology*, 104A:695 – 708.
- Steck, S. D. and Mallot, H. A. (2000). The role of global and local landmarks in virtual environment navigation. *Presence. Teleoperators and Virtual Environments*, 9:69 – 83.
- Veen, H. A. H. C. v., Distler, H. K., Braun, S. J., and Bühlhoff, H. H. (1998). Navigating through a virtual city: Using virtual reality technology to study human action and perception. *Future Generation Computer Systems*, 14:231 – 242.
- Wohlgemuth, S., Ronacher, R., and Wehner, R. (2001). Ant odometry in the third dimension. *Nature*, 411:795–798.

Granularity Transformations in Wayfinding

Sabine Timpf¹ and Werner Kuhn²

¹Department of Geography
University of Zurich
timpf@geo.unizh.ch

²Institute for Geoinformatics
University of Muenster
kuhn@ifgi.uni-muenster.de

Abstract. Wayfinding in road networks is a hierarchical process. It involves a sequence of tasks, starting with route planning, continuing with the extraction of wayfinding instructions, and leading to the actual driving. From one task level to the next, the relevant road network becomes more detailed. How does the wayfinding process change? Building on a previous, informal hierarchical highway navigation model and on graph granulation theory, we are working toward a theory of granularity transformations for wayfinding processes. The paper shows the first results: a formal ontology of wayfinding at the planning level and an informal model of granularity mappings.

Keywords: vehicle navigation, wayfinding, hierarchies, activity theory, graph granulation.

1 Introduction

Graph granulation theory [12] is neutral with respect to the choice of graph elements at a particular granularity level. This choice has to be guided by domain models, leading to application-specific network ontologies. A minimal ontology of road networks can be derived from a formalization of wayfinding activities at each level [9]. This idea is being applied here to a formalization of Timpf's hierarchical highway navigation process model [16].

Human beings use several conceptual models for different parts of geographic space to carry out a single navigation task. Different tasks require different models of space, often using different levels of detail. Each task is represented in a conceptual model and all models together form a cognitive map or collage for navigation. The different models of space need to be processed simultaneously or in succession to completely carry out a navigation task. Humans are very good at that type of reasoning and switch without great effort from one model of space to another. Today's computational navigation systems, on the other hand, cannot deal well with multiple representations and task mappings between them.

Existing spatial data hierarchies refine objects and operations from one level to the next, but the objects and operations essentially stay the same across the levels [1].

Task hierarchies, by contrast, refine the tasks from one level to the next and the objects and operations change with the level [7].

Timpf's cognitive architecture of interstate navigation [16] consists of three distinct conceptual models (levels): planning, instructing, and driving. Each level is characterized by a function computing information about a route. Each function takes the result of its predecessor and computes the route in the road network at its level. Thus, a concatenation of these wayfinding functions leads from the origin and destination of a trip all the way to detailed driving behavior.

The purpose of our work in progress is to gain a better understanding of these wayfinding functions and of the granularity mappings they induce on road networks. We present an executable formal specification of selected navigation functions in the functional language Haskell, specifically in its HUGS dialect [11]. Functional languages have great appeal for software engineering, because algebraic specifications [8] can be written and tested in them [6]. By the same token, they serve as test beds for formal algebraic theories.

The paper presents methodological and domain-specific results. Methodologically, we show that the use of functional languages for ontologies goes beyond collections of abstract data type specifications to comprehensive, layered object *and task* models with mappings among them. For the domain of highway navigation, we present the first (planning) level of a formal hierarchical task model.

The results are of interest to the spatial reasoning, geographic information science, and cognitive science communities. They touch on general issues in navigation and wayfinding, hierarchical reasoning, and formal ontological modeling. In practice, such formalizations of wayfinding can be used as algebraic specifications for more sophisticated navigation systems [17], supporting the planning, instructing, and driving processes on highway networks.

The remainder of the paper is structured as follows: section 2 presents the conceptual model for the three levels of wayfinding tasks; section 3 shows the granularity mappings among task levels by examples; section 5 presents the formalization approach; and section 6 discusses the results and future work.

2 The Conceptual Model

A series of complex cognitive operations are involved in planning and executing a journey on a highway network. Some operations are more general at one level (e.g., take exit) and are broken down into several operations at another level (e.g., change to appropriate lane for exiting, take ramp). There are operations that require bodily actions in one level but do not affect any other level (e.g., accelerating the vehicle). One of the authors has previously structured highway navigation operations into three task levels: planning, instructing, and driving (Timpf, et al. 1992). Table 1 shows how they organize the activity of wayfinding on highways and how they can be further subdivided into operations.

Table 1. Informal task ontology of wayfinding

Activity	Wayfinding: get from A to B		
Task Levels	<i>Plan</i> make a plan	<i>Instruct</i> produce instructions	<i>Drive</i> carry out instructions
Operations	find routes, determine constraints	take entrance, follow highway, change highway, take exit	take onRamp, change lane, change speed, proceed to, take offRamp

The planning, instructing, and driving tasks operate in different spatial domains. Planning involves knowledge about the places where one is and where one wants to go, as well as relevant additional places in between and the overall highway network containing them. The instruction task involves knowledge about the decision points along the route resulting from the planning task. The driving task needs information about when to drive where and how, but also introduces a body of actions of its own (e.g., change lane).

2.1 Ontologies at the Three Task Levels

At the Planning Level objects of the following types exist: *Place*, *Highway*, and *PLGraph* (the highway network at this level). The origin and destination of a trip are instances of places. Fig.1 shows an excerpt from the US highway network, labeled with place and highway names.

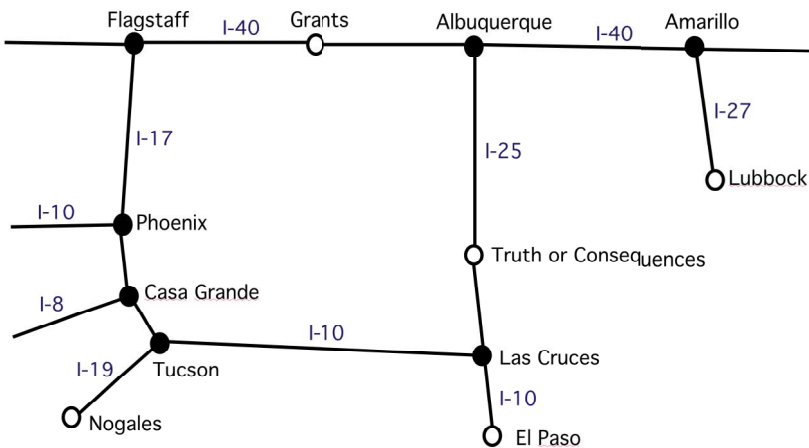


Fig. 1. Objects at the Planning Level (for a part of the US highway network located in New Mexico)

The *Instructional Level* (Fig. 2) introduces objects of type *Entrance*, *Exit*, *Section*, *Junction*, and *ILGraph* (the highway network at this level). A *Section* leads from an entrance to an exit on the same highway, while a *Junction* connects an exit to an entrance on another highway.

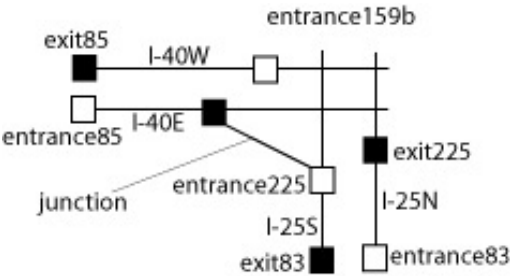


Fig. 2. Objects at the Instructional Level

The *Driving Level* (Fig. 3) is the most detailed, containing the objects and operations necessary to drive a vehicle with the instructions gotten at the previous level. Its pertinent objects are *lanes*, *ramps*, and the *DLGraph* (the highway network at this level). Three kinds of lanes exist: *travel*, *passing*, and *breakdown lanes*. *OnRamps* lead onto a highway, while *offRamps* leave a highway.

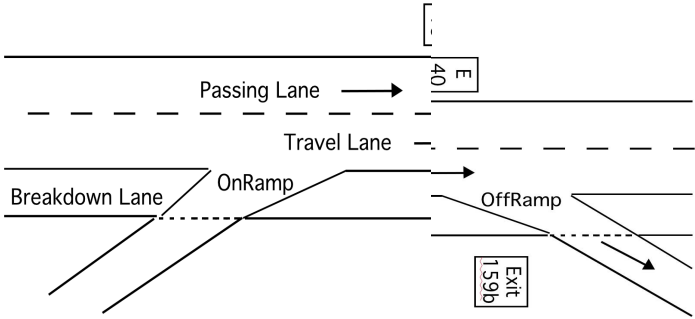


Fig. 3. Objects at the Driving Level

2.2 Graph Model

The objects at the Planning, Instructional, and Driving Levels are best represented by graphs and their parts (Fig. 4). The graph at the Planning Level contains places as nodes; highways are represented by named sequences of nodes connected by undirected edges. At the Instructional Level, nodes stand for exits and entrances, while directed edges represent highway sections and junctions. At the Driving Level, nodes represent ramps and (directed) edges represent lanes.

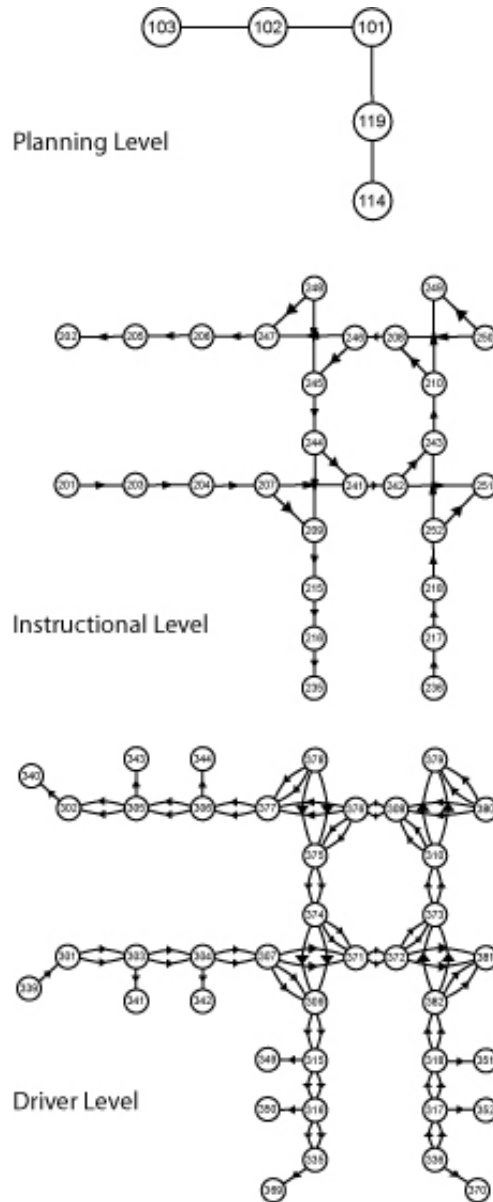


Fig. 4. Graphs representing the highway network at the three levels of detail

2.3 Reasoning in the Network: Navigation Tasks

Spatial reasoning in a highway network is a top-down process. Given the origin and destination of a trip on a highway network, reasoning at the Planning Level returns a *plan*. This plan is a list of places and highways necessary to get from the origin to the

destination, passing through highway interchanges. It is fed into the Instructional Level, which produces a list of *instructions*. These in turn are used as input to the Driving Level, which transforms them into driving *actions*.

The major operation at the Planning Level is to find a path from the origin to the destination. This path (which is a sequence of places to travel through) is then expressed as a sequence of place and highway names; e.g., (<Grants, I-40>, <Albuquerque, I-25>, <Truth or Consequences, reached>).

Table 2. Reasoning at the Planning Level

<i>plan (origin, destination, plGraph)</i>	
reasoning chain:	(<origin, HighwayName>, <Place, HighwayName>, ... <destination, "reached">)

The major operation at the *Instructional Level* (Table 3) is to produce instructions for a given plan. Information on the direction of the highway sections is taken from the highway network at this level, producing a sequence of triples <HighwayName, HighwayDirection, Distance>. The reasoning chain starts with finding the entrance, then taking it, and following the first highway to the first relevant interchange. This is repeated for all highways in the plan, followed by taking the exit at the destination place.

Table 3. Reasoning at the Instructional Level

<i>instructions (plan, ilGraph)</i>	
reasoning chain:	take_entrance (origin, firstHighway), follow <HighwayName, HighwayDirection, Distance> to interchange change_at junction, follow <HighwayName, HighwayDirection, Distance > to interchange ... take_exit at destination.

The operations at the *Driving Level* (Table 4) involve the actions to get from the *origin* to the *destination* with the help of the instructions. The *onRamp* brings one onto the acceleration lane, where it is necessary to *accelerate* and then to *change lane* to the *left* before being on the *highway*. Then, one *follows* the *highway* until the sign with the interchange mentioned in the instructions comes up and actions are required again. Then one has to *changeover* to the rightmost lane to be able to exit, an action composed of *decelerating* and taking the *offRamp*. In case of a *junction*, the driver will *proceed* to the next *highway* and accelerate again.

Table 4. Reasoning at the Diving Level

<i>drive (instructions, diGraph)</i>	
reasoning chain:	<pre> take_OnRamp (firstHighwaySection, firstDirection), accelerate, change_lane(left), follow < HighwayName, HighwayDirection, Distance> to (landmark), change_lane(right) until (rightneighbor(lane) = BreakdownLane), decelerate, take_OffRamp <HighwaySection, Direction>, proceed < HighwaySection, Direction>, accelerate, ..., take_OffRamp <HighwaySection, destination> </pre>

At this level, it is assumed that the driver knows how to steer a vehicle, how to accelerate or how to proceed. These actions are not further broken down. It is also assumed that the driver knows how to handle the car in the presence of other cars.

3 Granularity Mappings

Graph granulation theory posits two operations for graph simplification: *selection* and *amalgamation* (Stell and Worboys, 1999). Selection retains the nodes from one level that will be represented at a level of less detail. Any non-selected nodes disappear at the coarser level of detail. Amalgamation maps some paths at one level to nodes at a coarser level of detail.

Amalgamation is the simplification operation among our three levels of granulation. For example, in Fig. 4, the path leading from node 339 to 301 at the Driving Level is collapsed to node 201 at the Instructional Level. We have identified four different types of amalgamation in highway networks:

- Path -> Node
- Path -> (simplified)Path
- connected sub-graph -> node
- multi-Edge -> single Edge

The mappings between the complete graph at a level and the corresponding path as well as between paths and routes are selections (Fig. 5). The selection process leading from paths to routes (when each is seen as a graph) is exactly the selection operation of graph granulation theory. The selection process leading from the complete graphs to paths is a special selection operation producing a sub-graph of the original graph.

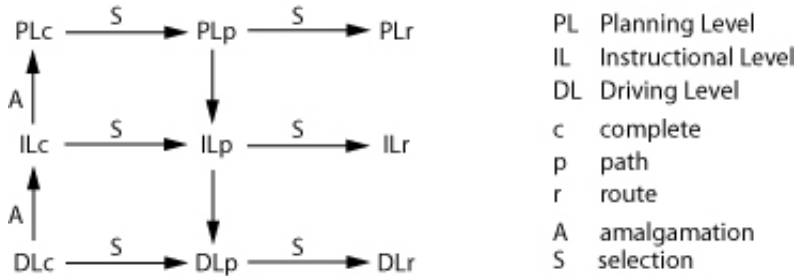


Fig. 5. Mappings

The correspondences between objects at different levels, resulting from the amalgamations, are shown in Table 5. They represent a special case of the aggregation and generalization hierarchies defined for graph objects in ([14]; [13]). Object types that are not explicit parts of the ontologies are put in parentheses (for example, individual edges do not play a role at the Planning Level).

Table 5. Corresponding objects

	Graph	Path	Edge	Node
Planning Level	PLGraph	Highway Route	(HwSegment)	Place (Interchange)
Instructional Level	ILGraph	Route	Section Junction	Exit Entrance
Driving Level	DLGraph	Route	Lane	OnRamp OffRamp

Our goal is a hierarchical graph structure that represents these amalgamations. Since the actual mappings are different for each instance (e.g., places can contain various combinations of exits and entrances, linked by sections and junctions; sections and junctions may consist of any number of lanes), this structure can only be described extensionally. Graph granulation theory (Stell and Worboys, 1999) proposes simplification graphs to represent the composition of each higher-level object from lower-level objects explicitly.

4 Formalization

We formalize our navigation model in HUGS, a dialect of the functional programming language Haskell [11]. Functional specifications serve as a workbench on which theories (e.g., of spatial cognition) can be

- worked out concisely and with formal algebraic rigor,
- tested for correctness (with respect to the requirements),
- adapted to ontological commitments

- compared and
- combined with each other [4].

Functional languages also have a great appeal for software engineering, because algebraic specifications [8] can be written and tested in them [6]. In this context, they combine the benefits of

- clean semantics (in particular, referential transparency for equational reasoning as well as a clean multiple inheritance concept),
- executability (allowing software engineers to test what they specify), and
- higher order capabilities (leading to leaner, more elegant descriptions).

Encouraged by a series of successful applications to non-trivial software engineering tasks ([2], [5], [15]), we have used functional languages for ontological research into the structure of application domains ([10]; Frank 1999; Kuhn 2001). The work presented here continues on this path by formalizing hierarchical navigation tasks on highway networks.

The object classes of the data model (i.e., the Haskell data types) are based on notions of graph theory. For instance, a highway section is an edge between nodes in the highway network at the Instructional Level. We are using Erwig's inductive graph library [3] to supply the necessary graph data types and algorithms. The HUGS code below also uses some elementary list functions. Parentheses construct tuples (specifically, pairs), and brackets construct lists.

The data type definitions formalize the ontologies for each task level:

```
-- Planning Level object types
type Place = Node
type Highway = (HighwayName, [Place])
type PLGraph = Graph PlaceName EdgeLength
type Route = Path
type Plan = [(PlaceName, HighwayName)]

-- Planning Level attributes and auxiliary types
type PlaceName = Name
type HighwayName = Name
type Highways = [Highway]
type Leg = (Place, Highway)
type Legs = [Leg]

-- Instruction Level object types
type Entrance = Node
type Exit = Node
type Section = Edge
type Junction = Edge
type ILGraph = Graph EName EdgeLength

-- Instruction Level attributes and auxiliary types
type EName = Name

-- Driving Level object types
```

```

type Ramp = Edge
type Lane = Edge
type DLGraph = Graph Name EdgeLength

```

At the planning level, the route from origin to destination is determined by the shortest path operation (sp) applied to the highway graph at this level (PLGraph).

```

route :: Place -> Place -> PLGraph -> Route
route origin destination plg = sp origin destination plg

```

This route is a path in the graph, i.e., a sequence of nodes. It has to be translated into a plan, i.e., a sequence of pairs with names of places and highways to follow. For this purpose, information about the highways has to be combined with the route and the graph. This is done in a function computing the “legs” (a sequence of pairs of places and highways) leading from origin to destination:

```

legs :: Route -> Highways -> Legs
legs (x:[]) hws = [(x, endHighway)]
legs rt hws = (head rt, firstHighway rt hws) : legs (tail rt)
hws

```

The recursively applied function firstHighway computes the first highway to take on a route:

```

firstHighway :: Route -> Highways -> Highway
firstHighway rt hws = fromJust (find (hwConnects (rt !! 0)
(rt !! 1)) hws)

```

The first highway is determined by finding, among all highways, the highway that connects the first and second place (assuming there is only one):

```

hwConnects :: Place -> Place -> Highway -> Bool
hwConnects p1 p2 hw = (elem p1 (snd hw)) && (elem p2 (snd
hw))

```

From the legs of the trip, those legs which continue on the same highway can be eliminated:

```

planModel :: Legs -> Legs
planModel lgs = map head (groupBy sameHighway lgs)

sameHighway :: Leg -> Leg -> Bool
sameHighway (p1, hw1) (p2, hw2) = hw1 == hw2

```

Finally, this (internal) model of a plan is translated into an (external) view expressing it by the names of places and interchanges:

```

planView :: Legs -> PLGraph -> Plan
planView (x:[]) plg = [(placeName (fst x) plg, fst (snd x))]
planView pm plg = (placeName (fst (head pm)) plg, fst (snd
(head pm))) : (planView (tail pm) plg)

```

This completes the formal model at the Planning Level. The given HUGS code allows for the computation of trip plans on any highway network that is expressed as an inductive graph.

At the Instructional Level, the planModel will get expanded into a list of highway entrances, segments, junctions, and exits, using the amalgamation functions to be

defined. Similarly, at the Driving Level, these instructions will be expanded into driving actions consisting of ramps and lanes to take.

5 Conclusions

Human beings use information at multiple levels of detail when navigating highway networks. This paper describes a conceptual model of the U.S. Interstate Network at three levels of reasoning: planning, instructing, and driving. The apparently simple everyday problem of navigating a highway network has been shown to contain a high degree of structure and complexity. Executable algebraic specifications and graph granulation theory have been applied to formalize this structure and test the results.

The formalization presented in this paper covers the first level of reasoning (planning tasks). It provides a framework for comparing the reasoning at the three levels. While planning involves the computation of a shortest path, finding instructions and transforming them to driving actions use granulation relationships between graphs, rather than graph operations at a single level. The definition of and interaction between the three levels is intended to provide a cognitively plausible model of actual human wayfinding processes within the U.S. Interstate Highway Network. We proposed objects and actions corresponding to the *physical* structure at each level and playing a role in real wayfinding processes.

The formal model can serve as a software specification (specifically, as the essential and abstract model) for navigation systems used for Interstate travel. Software for navigation systems is currently very limited in its support for hierarchical reasoning. The key benefits of choosing a functional language to write algebraic specifications for navigation operations are that the specified models can be tested and are semantically unambiguous.

Acknowledgments

The work reported here was supported by the University of Zürich, the University of Münster, and the Technical University of Vienna.

References

- [1] Car, A. (1997). Hierarchical Spatial Reasoning: Theoretical Consideration and its Application to Modeling Wayfinding. GeoInfo Series Vol. 10. TU Vienna: Dept. of Geoinformation.
- [2] Car, A. and A. U. Frank (1995). Formalization of Conceptual Models for GIS using Gofer. *Computers, Environment, and Urban Systems* 19(2): 89-98.
- [3] Erwig, M. (2001). Inductive Graphs and Functional Graph Algorithms. *Journal for Functional Programming* 11(5): 467-492.
- [4] Frank, A. U. (1999). One step up the abstraction ladder: Combining algebras - From functional pieces to a whole. *Spatial Information Theory*. C. Freksa and D. Mark, Springer-Verlag. Lecture Notes in Computer Science 1661.

- [5] Frank, A. U. and W. Kuhn (1995). Specifying Open GIS with Functional Languages. *Advances in Spatial Databases - 4th Internat. Symposium on Large Spatial Databases, SSD'95 (Portland, ME)*. M. Egenhofer and J. Herring. New York, Springer-Verlag: 184-195.
- [6] Frank, A. U. and W. Kuhn (1999). A Specification Language for Interoperable GIS. *Interoperating Geographic Information Systems*. M. F. Goodchild et al., Kluwer: 123-132.
- [7] Freksa, C. (1991). Qualitative Spatial Reasoning. In D. M. Mark & A. U. Frank (Eds.), *Cognitive and Linguistic Aspects of Geographic Space*. Dordrecht, The Netherlands: Kluwer Academic Press: 361-372.
- [8] Guttag, J. V. (1977). Abstract Data Types and the Development of Data Structures. *ACM Communications* 20(6): 396-404.
- [9] Kuhn, W., 2001. Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, 15(7): 613-631.
- [10] Medak, D. (1997). Lifestyles - A Formal Model. Chorochronos Intensive Workshop '97, Petronell-Carnuntum, Austria, Dept. of Geoinformation, TU Vienna.
- [11] Peterson, J., K. Hammond, et al. (1997). The Haskell 1.4 Report. <http://haskell.org/report/index.html>.
- [12] Stell, J. G., & Worboys, M. F. (1999). Generalizing Graphs using amalgamation and selection. In R. H. Gueting & D. Papadias & F. Lochovsky (Eds.), *Advances in Spatial Databases, 6th Symposium, SSD'99* (Vol. 1651 LNCS, pp. 19-32): Springer.
- [13] Timpf, S. (1999). Abstraction, levels of detail, and hierarchies in map series. *Spatial Information Theory - cognitive and computational foundations of geographic information science*. C. Freksa and D.M. Mark. Berlin-Heidelberg, Springer-Verlag. Lecture Notes in Computer Science 1661: 125-140.
- [14] Timpf, S. (1998). Hierarchical structures in map series. GeoInfo Series Vol. 13. Vienna: Technical University Vienna.
- [15] Timpf, S. and A. U. Frank (1997). Using Hierarchical Spatial Data Structures for Hierarchical Spatial Reasoning. *Spatial Information Theory - A Theoretical Basis for GIS (International Conference COSIT'97)*. S. C. Hirtle and A. U. Frank. Berlin-Heidelberg, Springer-Verlag. Lecture Notes in Computer Science 1329: 69-83.
- [16] Timpf, S., G. S. Volta, et al. (1992). A Conceptual Model of Wayfinding Using Multiple Levels of Abstractions. *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*. A. U. Frank, I. Campari and U. Formentini. Lecture Notes in Computer Science 639: 348-367.
- [17] White, M. (1991). Car navigation systems. *Geographical Information Systems: principles and applications*. D. J. Maguire, M. F. Goodchild and D. W. Rhind. Essex, Longman Scientific & Technical. 2: 115-125.

A Geometric Agent Following Route Instructions*

Ladina B. Tschander, Hedda R. Schmidtke, Carola Eschenbach,
Christopher Habel, and Lars Kulik

University of Hamburg, Department for Informatics
Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
{tschander, schmidtke, eschenbach, habel, kulik}
@informatik.uni-hamburg.de

Abstract. We present the model of a Geometric Agent that can navigate on routes in a virtual planar environment according to natural-language instructions presented in advance. The Geometric Agent provides a new method to study the interaction between the spatial information given in route instructions and the spatial information gained from perception. Perception and action of the Geometric Agent are simulated. Therefore, the influence of differences in both linguistic and perceptual skills can be subject to further studies employing the Geometric Agent. The goal of this investigation is to build a formal framework that can demonstrate the performance of specific theories of the interpretation of natural-language in the presence of sensing. In this article, we describe the main sub-tasks of instructed navigation and the internal representations the Geometric Agent builds up in order to carry them out.

1 Introduction

When humans have to solve the problem ‘How to come from A to B’ in an unknown environment, querying for a verbal route instruction can be helpful. Formulated in a more general way, communication about space can facilitate spatial problem solving. The overall criterion for the adequacy of a route instruction is whether it enables navigators to find their way. Thus, adequacy depends on a wide spectrum of parameters. For example, epistemological parameters, such as the knowledge of the participants (the instructor and the instructee), or perceptual parameters, which concern the navigator’s perception of the environment and the perceptual salience of landmarks, can influence the performance of the navigator. Crucial linguistic parameters range from the modus of the utterance, e.g. declarative vs. imperative, to the type and quantity of the spatial information provided by the route description.

* The research reported in this article was supported by the Deutsche Forschungsgemeinschaft (DFG) and carried out in the context of the project “Axiomatik räumlicher Konzepte” (Ha 1237-7) that is imbedded in the priority program on “Spatial Cognition”. We thank the participants of the “route instruction” project (academic year 2001/02) for support in the collection of verbal data and the analysis of navigation tasks, and two anonymous reviewers for helpful comments and suggestions.

In this article, we discuss navigation based on instructions given in advance. In particular, we focus on cases where neither the instructor nor the instructee perceives the environment, the critical landmarks, or the tracks (roads, footpaths, trails).¹ During the instruction phase, an instructor who is familiar with the environment in question produces a route instruction. In comprehending the instruction, the instructee builds up conceptual, mental representations of the route. These representations, which contain spatial information about the route and the sequence of actions to be performed, have to be stored in memory. In the navigation phase, the instructed navigator has to match the internal representations against the perceived scenes. This process involves the recognition of spatial configurations of landmarks, tracks, and positions in accordance with the spatial relations specified in the instruction. Route instructions normally do not specify all objects or spatial configurations that the navigator will perceive on the route (Tversky & Lee 1999). Therefore, the spatial representation of the route is enriched during the navigation phase, and the communicated sequence of actions has to be refined.

Starting with section 3, we discuss processes and components that are involved in the interaction between spatial information given in route instructions and spatial information gained from perceiving the environment during navigation. We propose the idea of a *Geometric Agent* that simulates instructed navigation in a virtual planar environment.² The Geometric Agent serves as a formal framework for testing specific proposals for solving the individual tasks and for the interpretation of route instructions. It will yield an operational framework for testing the adequacy of route instructions for navigators whose information processing capabilities, linguistic and spatial knowledge are completely known and can be subject to modifications.

In our approach, the conceptual representation of a route instruction is encoded in the Conceptual Route Instruction Language CRIL. CRIL-expressions are constructed from a basic inventory of descriptive operators (see section 3). On the one hand, CRIL-expressions specify the semantics of natural language expressions in the traditional method of formal semantics. On the other hand, CRIL is an internal language of the Geometric Agent that relates to perceptual objects and specifies actions the Geometric Agent can carry out. CRIL and formal reasoning based on CRIL-expressions can be used to test contrasting proposals for the semantics of spatial expressions regarding their consequences for the performance of an instructed navigator. Additionally, CRIL can be seen as an Agent Communication Language (see Labrou et al., 1999), i.e., CRIL provides artificial agents with a means to communicate, in particular to exchange knowledge about routes.

¹ Most authors count the tracks among the landmarks (for example Allen 1997, Denis 1997, or Lovelace et al. 1999). Tracks can function both as (local) landmarks helping to identify a position on the route and as guiding structures for low-level navigation.

² The geometric agent proposed in the present paper is kindred to *Homer* (Vere & Bickmore 1990), *BEELINE* (Mann 1996) and the idea of the map-making agent and the map-using agent (Frank 2000).

A prototypical realization of basic components is available via
<http://www.informatik.uni-hamburg.de/WSV/Axiomatik-english.html>

In the model of the Geometric Agent, the conceptual representation of the route instruction separates the spatial information and the action plan. The spatial information of the route description is represented as a net-like structure – the CRIL-net – that abstracts from linguistic details of the route description. The action plan constitutes a sequence of commands, which employ a small set of imperative operators and refer to nodes in the CRIL-net. Imperative operators describe desired actions and states, i.e., they have a declarative—not a procedural—character (Labrou et al., 1999). According to the plan-as-communication view (Agre & Chapman 1990), the Geometric Agent interprets the imperative operators dependent on the situation.

During the navigation phase, evaluating spatial relations in the perceived scene is a multi-modal task. Thus, the Geometric Agent provides a framework for testing theories on the interaction between propositional and non-propositional representations in instructed navigation. Landmark information characterizes objects that the navigator has to look for. Therefore, the usefulness of landmark information relates to the perceptual abilities of the navigator. The goal of our investigations is to study the influence of the quality of information gained from perception on the usefulness of a route description rather than the task of landmark recognition. Therefore, the perception of the Geometric Agent is simulated such that its ability to recognize and identify landmarks from different positions can be completely controlled.

The outline of the article is as follows: In the next section, we review the characteristics of verbal route instructions and give two examples that escort the following discussion. In the third section, we discuss the instruction phase and the sources of information contributing to the internal model of the route. The fourth section presents the Geometric Agent's interaction with the geometric environment. The final section discusses the tasks to be performed in the navigation phase.

2 Route Instructions³

Route instructions specify spatial information about the environment of the route and temporal information about the actions (movements, turns) to be performed (Denis, 1997). Information about routes can be communicated in different ways. Natural language descriptions are a typical means. Routes can also be presented with a list of written notes about relevant actions, they can be depicted as strip maps, or they can be marked by a salient line on a map. Furthermore, different modalities can be combined. For example, a navigational assistance system can combine a verbal description of an action with the display of an arrow that indicates a direction. In face-to-face instructions, verbal descriptions are regularly supported by gestures. In this section, we describe which information can be extracted from a mono-modal, verbal route instruction given in advance.

³ The terms “route description”, “route instruction”, and “route direction” have been used to refer to the same type of discourse conveying information about routes. Since the modus of individual utterances (declarative vs. imperative) is not in the focus of our discussion, we use the terms “route instruction” and “route description” interchangeably in this article.

2.1 Verbal Route Instructions

For more than twenty years, route instructions have been subject to interdisciplinary research in linguistics and psychology (Klein, 1979, 1982; Wunderlich & Reinelt, 1982; Allen, 1997; Denis, 1997). These investigations focus on the communicative role of the instructor and the task of producing the route description. The main sub-topics are: the overall discourse situation of route instructions, the structure and content of the texts, the linguistic items employed, and the relation between the spatial knowledge of the instructor and the task of generating a linguistic description.⁴

There is a strong agreement regarding the tasks to be solved during producing a route description. Instructors have to activate a mental representation of an area containing the starting position and the goal. Then, they select a suitable route connecting the starting position with the goal. Furthermore, instructors have to decide which objects of the environment can function as landmarks. Finally, they have to produce the verbal description.

In contrast to the production perspective, this article focuses on the role of the instructee and the interpretation of route instructions in relation to the conceptual representations gained from perception during navigation. According to production models, we assume different types of representation of route knowledge. The instructee transforms the verbal route instruction into a conceptual representation of the route that connects the starting point with the goal. In addition, the instructee extracts an action plan that consists of a sequence of situation-sensitive instructions representing temporal aspects of the route. During the navigation phase, more detailed information about spatial and temporal aspects of the route can be added.

In comprehending the instruction, the instructee builds up a representation of the sentence meaning based on linguistic knowledge (syntax, lexicon). This representation has a net-like structure rather than a map-like structure (see Werner et al., 2000). From this representation of the route, spatial information is extracted and gaps in the route can be closed via inferences. Since the instructee does neither know nor perceive the environment, the resulting spatial representation is underdetermined regarding distances between landmarks and angles between tracks.

Wunderlich and Reinelt (1982) deal with the structure of route instructions in German from a discourse theoretic perspective. They distinguish three types of segments of the route. The starting segment of a route contains the starting point and the initial orientation of the navigator. The middle part of the route instruction consists of a sequence of intermediate segments, which can be specified by the designation of landmarks, the reorientation of the navigator, the start of a progression, and its end. The intermediate segments are linguistically combined with *und dann* [*and then*], *danach* [*after that*], or *bevor* [*before*]. *Bis* [*until*] marks the end of an intermediate

⁴ The cited examinations form the basis of further investigations of route descriptions. For example, there is research on criteria for 'good' route descriptions (Lovelace, Hegarty & Montello 1999), on the influence of the kind of environment on the description of the route (Fontaine & Denis 1999), on comparing the structure of depictions and descriptions of routes (Tversky & Lee 1999), on generating cognitive maps based on linguistic route descriptions (Fraczak 1998), and on the generation of linguistic route descriptions (Ligozat 2000).

segment. The use of *direkt* [directly] or *genau* [exactly] indicates the final segment including the goal as a perceivable object.

Although verbal route instructions exhibit a great variability, the general structure of route instructions—as described by Wunderlich and Reinelt—seems to be quite common. This general organization of the discourse structure of route instructions can be used for extracting the spatial information of the route and for closing gaps in individual route instructions.

Klein (1979), Wunderlich and Reinelt (1982), Allen (1997), and Denis (1997) agree that two types of information are prominent in route descriptions. On the one hand, information about landmarks and decision points, and, on the other hand, information about actions the navigator has to perform. Decision points, which are positions on the route on which the navigator can choose between different tracks, are mostly characterized in their relation to landmarks. The importance of landmarks and their role of marking decision points is confirmed by many psychological studies (among others Tversky, 1996; Denis, 1997; Allen, 1997; Fontaine & Denis, 1999; Tversky & Lee, 1999). However, landmarks can also be given along a longer track assuring the navigator is still on the right track (Lovelace, Hegarty & Montello, 1999). The order in which decision points and landmarks appear in an instruction is organized according to a virtual navigator (Klein, 1979, used the German term “imaginärer Wanderer”). The virtual navigator provides a reference system, which can be used for grounding projective relations.

The primary actions named in route instructions are movements and changes of orientation. Denis (1997) adds positioning as a third kind of prescriptions that occur in route instructions. In the internal models of the Geometric Agent, the operators !GO, !CH_ORIENT and !BE_AT represent instructions to perform the actions that can be described by verbs as *go*, *turn*, and *be*. Allen (1997) describes these verbs as typical indicators for the three types of prescriptions.⁵

2.2 Simple Route Instructions: Two Examples

We collected eight instructions of a route between two buildings on the campus of the Department for Informatics of the University of Hamburg. All the informants know the campus well. They were orally asked to describe the route from the dining hall (in house ‘B’) to house ‘E’ for a person that does not know the campus (see Figure 1). The informants produced written descriptions of the route from memory. This took place inside house ‘F’, i.e. spatially detached from the route to be described. All descriptions contain the three segments identified by Wunderlich and Reinelt (1982) and Denis (1997).

⁵ Our approach to the semantics of route instructions (see section 3) follows the line developed by Crangle and Suppes (1994). Their model-theoretic approach for the semantics of commands requires that conditions of satisfactory execution are represented in addition to the procedures that execute an instruction. Comparable to this, the representation of the spatial information can specify conditions that have to be fulfilled after executing an action.

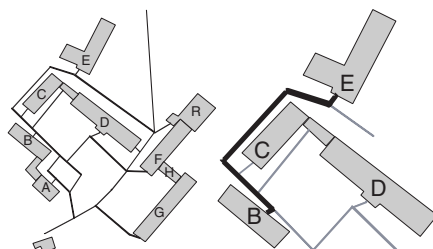


Fig. 1. The buildings and tracks of the campus of the Department for Informatics of the University of Hamburg. The route described in the examples (1) and (2) is indicated by a thick black line

Two of these texts serve to illustrate the following discussion. Instruction (1) is formulated in declarative mode. The indefinite pronoun *man* [one] is used to refer to the navigator. Several landmarks are mentioned, such as houses, tracks, a gate, a fence, and a square. In instruction (2) the imperative mode is used in the main clauses of the intermediate segments. In subordinate clauses and in the last sentence (final segment), the (informal) personal pronoun *du* [you] refers to the navigator. Tracks are not mentioned and houses are the only type of landmarks used in this instruction.

(1) (a) Um von der Mensa zum Haus E zu gelangen, [for reaching house E from the dining hall] (b) hält man sich nach dem Verlassen der Mensa links [one keeps left after leaving the dining hall] (c) und geht auf die geschlossene Pforte zu. [and walks towards the closed gate] (d) Auf diesem Weg trifft man auf eine Abzweigung eines kleinen Weges nach rechts, [on this track one meets an junction with a small track to the right] (e) der zwischen Zaun und Haus entlangführt. [that leads along between fence and house] (f) Dieser Weg mündet hinter dem Haus auf einem gepflasterten Platz, [the track leads behind the house on[to] a paved square] (g) von dem man mit einer Treppe in Haus E gelangt. [from which one reaches the house with stairs]	(2) (a) Wenn du aus der Mensa kommst, [when you leave the dining hall] (b) geh nach links, [walk to the left] (c) zwischen Haus B und Haus C durch. [through [the region] between house B and house C] (d) Geh hinter Haus C lang, [walk along behind house C] (e) und dann, wenn du an Haus C vorbeibist, [and then, when you are past house C] (f) wieder nach rechts. [again to the right] (g) Dann stehst du vor Haus E. [then you will stand in front of house E]

The two instructions are similar regarding the spatial information. The introduction (1a) summarizes the task to be performed and mentions the starting position and the goal. (1b) as well as (2a) refer to the starting position via the given landmark (dining hall) and give information about the initial orientation of the navigator by mentioning a movement (leaving the dining hall) the navigator can carry out. The integration of position and orientation—in robotics often called “pose”—is fundamental for navigation. For example, (1b) and (2b) specify the direction of the second movement relative to the initial pose (to the left). (1c) specifies the next movement as directed to a closed gate. (2c) specifies the same movement as crossing the region between two houses. (1d–f) describe the spatial constellations of the tracks to be perceived. The movements to be performed are not explicitly mentioned. In contrast to this, (2d) expresses a movement, but does not describe the junction of the tracks. In interpreting (2f), the instructee can infer from the particle *wieder* [again] that the right turn mentioned is not the first turn of its kind. Thus, it can be concluded that the two movements described by (2c) and (2d) are also connected by a right turn. (2f) describes the final right turn, which is not mentioned in the first text. (1g) and (2f) complete the description by expressing that the navigator reaches the goal via the stairs or the current location is in the front-region of the goal.

3 The Instruction Phase

When a route description is given in advance, the instruction phase is temporally separated from the navigation phase. The interpretation of a route description during the instruction phase is exclusively based on linguistic knowledge (grammar and lexicon) and on general knowledge about temporal and spatial concepts. Information about the spatial layout of the environment is not accessible until the navigation phase. The result of comprehending a route instruction is an internal representation of the route, the so-called “instruction model”. The transformation from a verbal instruction to its internal counterpart is modeled in the Geometric Agent by a two-step process (see Figure 2).

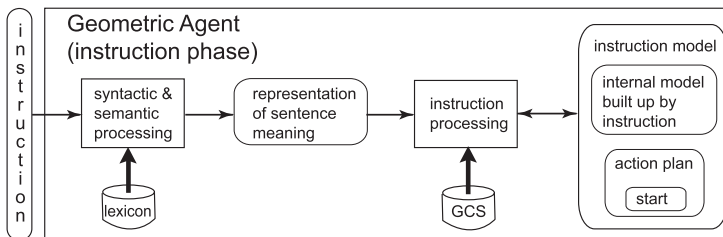


Fig. 2. Tasks in the instruction phase of the Geometric Agent

Firstly, verbal route instructions are rendered into “representations of the sentence meaning” combining the lexical entries according to the syntactic structure. These

representations contain spatial information about the route as well as temporal and ordering information about the sequence of actions to be carried out by the navigator. The component called “instruction processing” separates these two types of information. The spatial portion is used to construct an “internal model of the route”. This representation constitutes the core of the instruction model of the Geometric Agent. A second component, called the “action plan”, consists of a sequence of imperative statements. It specifies which actions have to be executed in which order. Both types of internal representations—constituting the instructive part and the descriptive part of the instruction model—are specified in the conceptual route instruction language CRIL. The declarative portion of CRIL is based on linguistic analyses of spatial expressions we described in Eschenbach et al. (2000) and Schmidtke et al. (to appear).

In the present section, we focus on the spatial information mediated by verbal route instructions in the instruction phase. Thus, we concentrate on aspects that depend on spatial knowledge rather than discuss general aspects of syntactic and semantic processing or of the representation of sentence or text meaning. Two modules containing spatial knowledge, namely the “Spatial Lexicon” (see section 3.1) and the “Geometric Concept Specification” (GCS; see section 3.5), play a major role in the construction of the instruction model.

3.1 The Contribution of the Spatial Lexicon

The lexicon is a module of linguistic knowledge that maps words onto structures representing their meaning. It combines syntactic and semantic information about the words such that the syntactic structure can support the derivation of the meaning of phrases and sentences. Thus, the task to construct the meaning of a route instruction presupposes a coherent and consistent system of entries in the spatial lexicon. The following proposal for entries of the spatial lexicon is based on linguistic analyses of spatial expressions (Jackendoff, 1990; Kaufmann, 1995; Eschenbach et al., 2000; Schmidtke et al., to appear). Our approach to the spatial lexicon uses axiomatic characterizations based on an inventory of “descriptive operators” to specify the semantic part of lexical entries (Eschenbach et al., 2000; Schmidtke et al., to appear). Table 1 lists some descriptive operators used in lexical entries discussed in the following.

Route instructions specify actions, paths, tracks, positions and landmarks in relation to each other. Different groups of words characterize these components. For example, the actions mentioned in route instructions are specifically described with verbs of position, verbs of locomotion, and verbs of change of orientation.

Verbs of position (e.g., *stehen* [*stand*]) include the component $BE\_AT(x, p)$ which expresses that object x is at position p .⁶ The semantic component $GO(x, w)$ is characteristic for verbs of motion (*gehen* [*go/walk*], *betreten* [*enter*], *verlassen* [*leave*]). It indicates that x moves along the path w . Verbs of change of orientation (*abbiegen*

⁶ The variable x stands for the bearer of action. Variables beginning with l are used for perceptible spatial entities such as landmarks and tracks. Variables for positions start with p , variables for paths with w , and variables for directions with d .

[turn off]) contain $CH\_ORIENT(x, d)$ representing the change of the orientation of x such that after the movement x is directed according to d (see Schmidtke et al., to appear). In route descriptions, the manner of motion is usually in the background. Correspondingly, verbs that are (more or less) neutral regarding the manner of motion occur (e.g., *gehen* [go/walk], *laufen* [run/walk], *fahren* [drive/ride], *folgen* [follow], *sich halten* [keep], *abbiegen* [turn off], *nehmen* [take]). The Geometric Agent does not consider the manner of motion specified by a verb, since it is not able to move in different ways.

Table 1. Descriptive operators used in lexical entries of verbs or prepositions⁷

type of natural language expression	characteristic semantic component
verbs of position	$BE\_AT(x, p)$
verbs of motion	$GO(x, w)$
verbs of change of orientation	$CH\_ORIENT(x, d)$
local prepositions or adverb	$LOC(u, PREP(l))$
directional preposition or adverb	$TO(w, PREP(l))$ $FROM(w, PREP(l))$ $VIA(w, PREP(l))$ $LOC(w, PREP(l))$
projective terms	$PREP(l, rsys)$

Tracks are naturally or artificially marked elongated structures in the environment. Paths of motion in the instruction model represent trajectories of objects in motion. Paths are directed and not perceivable in a static situation. They are introduced by natural language expressions that distinguish a starting point and a final point. Due to their different spatial properties, paths and tracks cannot be identified, but paths can run along tracks and tracks can originate from people, animals or vehicles moving along the same paths.

Paths of motion are characterized by verbs of motion and by directional prepositional phrases and directional adverbs. Complex adverbial phrases can describe (complex) configurations of paths, as in example (2d–f) above. In addition, the description of tracks can implicitly specify paths of motion as in (1e). Thus, in route instructions a path can be introduced by a phrase that does not explicitly specify the corresponding action. Correspondingly, actions that are not mentioned explicitly in the route description can be explicitly introduced into the action plan.

Decision points on the route are mostly endpoints of paths. The spatial information of several directional prepositional phrases and adverbs concerns the position of the endpoints of the path. For example, the preposition *zu* [to], *nach* [to], and the directional versions of *an* [at], *auf* [on], *in* [in] etc. specify a region relative to a landmark

⁷ PREP is a place holder for a lexeme-specific function that maps landmark l to a spatial region. For example, the local preposition *in* [+Dat] is represented as $LOC(u, IN(l))$. The directional preposition *in* [+Akk] is represented as $TO(w, IN(l))$ and the directional preposition *aus* [+Akk] is represented as $FROM(w, IN(l))$. The semantic components LOC, TO, FROM, VIA, IN etc. are specified in the geometric concept specification GCS (see Eschenbach et al., 2000).

(PREP(l)) and express that the final point of the path is enclosed in the region and that the starting point is not enclosed (TO(w , PREP(l))). The prepositions *von* [from] and *aus* [out of] specify a region which encloses the starting point but not the final point (FROM(w , PREP(l))). The preposition *durch* [through] indicates a region that encloses an inner point of the path but not the starting point or final point (VIA(w , PREP(l))). Further information about regions can be given by local prepositional phrases that specify positions, decision points, or locations of landmarks relative to each other. Projective terms implicitly refer to a spatial reference system ($rsys$) that has to be anchored relative to the conceptual representation of the preceding segments of the route instruction (Klein, 1979, 1982; Levinson, 1996; Eschenbach, 1999).

Noun phrases such as *die Mensa* [the dining hall], *das Haus* [the house], *die Pforte* [the gate], and *der Zaun* [the fence] refer to landmarks. They combine with local or directional prepositions to specify regions including paths, positions, or decision points. Nouns such as *Weg* [track], *Straße* [street/road], and *Kreuzung* [crossing], *Abzweigung* [junction] relate to different types of tracks, or configurations of tracks. Tracks can function as landmarks or as specifying a path of motion. During the instruction phase, the function of the tracks mentioned has to be inferred.

3.2 Instruction Processing

In this section we exemplify the task of instruction processing with a discussion of instruction (2). Table 2 displays the three types of information that can be extracted from the route description: actions, spatial relations, and landmarks.

Table 2. CRIL-representation of example (2)

(2)		actions	spatial relations	landmarks
(a)	Wenn du aus der Mensa kommst, [when you leave the dining hall]		FROM(w_1 , IN(l_1))	MENSA(l_1)
(b)	geh nach links, [walk to the left]	!GO(w_2)	TO(w_2 , LEFT($rsys_2$))	
(c)	zwischen Haus B und Haus C durch. [through [the region] between house B and house C]	!GO(w_3)	VIA(w_3 , BETWEEN(l_2 , l_3))	HOUSE(l_2) NAME(l_2 , 'B') HOUSE(l_3) NAME(l_3 , 'C')
(d)	Geh hinter Haus C lang, [walk along behind house C]	!GO(w_4)	LOC(w_4 , BEHIND(l_3 , $rsys_4$)) ALONG(w_4 , l_3)	HOUSE(l_3) NAME(l_3 , 'C')
(e)	und dann, wenn du an Haus C vorbei bist, [and then, when you are past house C]	!BE_AT(p_1)	LOC(p_1 , PAST(l_3 , $rsys_5$))	HOUSE(l_3) NAME(l_3 , 'C')
(f)	wieder nach rechts. [again to the right]	!GO(w_6)	TO(w_6 , RIGHT($rsys_6$))	
(g)	Dann stehst du vor Haus E. [then you will stand in front of house E]	!BE_AT(p_2)	LOC(p_2 , FRONT(l_4 , $rsys_7$))	HOUSE(l_4) NAME(l_4 , 'E')

Actions are represented in CRIL by imperative statements of the form $!GO(w)$ [move along path w], $!BE\_AT(p)$ [verify that you are at position p , if not, move to p], $!CH\_ORIENT(d)$ [turn to direction d]. According to the view that imperative statements refer to desired actions and states, the “imperative operators” $!GO(w)$, $!BE\_AT(p)$, $!CH\_ORIENT(p)$ possess descriptive counterparts, namely the operators: GO , BE_AT , CH_ORIENT . For example, if navigator x follows the imperative statement $!GO(w)$ successfully in a situation s , then the descriptive statement $OCC(s, GO(x, w))$ is true.

Landmark information specifies objects that the navigator will perceive on the route. Based on the landmarks, regions are specified that include decision points or other positions of interest during the navigation phase. This type of spatial information is given in Table 2 under the heading “spatial relations”. For example, the expression $VIA(w_3, BETWEEN(l_2, l_3))$ relates the path w_3 , and the region between two landmarks l_2 and l_3 .

The information about the sequence of actions constitutes the action plan. The internal model of the route includes the spatial information and the landmark specifications. The internal model is enriched in two steps. Firstly, the instructee can employ the specification of the spatial concepts and pragmatic principles in the instruction phase. This type of inference is discussed in the remainder of this section. Secondly, the navigator can add information gained by experience in the navigation phase. Such additions are treated in section 4.

3.3 The Action Plan

The action plan is a list of imperative statements ($!GO(w)$, $!BE\_AT(p)$, $!CH\_ORIENT(p)$), which are interpreted during the navigation phase resulting in actions, in particular movements (see section 5). Since the imperative operators correspond to the descriptive operators supplied by verbs of motion, the sequence of imperative operators is given mainly by the verbs of the route instruction. The instruction processing extracts the corresponding imperative operators from the lexical entries of the verbs and lists them in the action plan according to their appearance in the route instruction. If the route instruction contains verbs that describe constellations of tracks (as in example (1d–f) given in section 2.2), the list of imperative operators is derived from the spatial relations (see section 3.6). In both cases, the sequential ordering of the imperative statements represents the temporal aspect of the route instruction. The action plan also specifies the starting pose, i.e., the navigator’s position and orientation at the beginning of the navigation phase.

3.4 The Internal Model Built up by Instruction

The internal model of the route is represented as a CRIL-net. CRIL-nets include nodes that represent landmarks (l), paths (w), tracks (t), regions (r), and positions (p). In verbal route instructions, landmarks can be described with non-spatial characteristics (as to which category a landmark belongs to or if it has a specific color). This kind of information is represented as attributes in CRIL-nets. Reference systems

(*rsys*) are included as nodes that create a demand to anchor the projective relation to the context. CRIL-nets of route instructions are related to route graphs (Werner et al., 2000), which are assumed to be acquired by navigation experience.

The different types of nodes are connected by labeled edges describing the spatial relations that hold between them. For example, region nodes are related to landmarks or reference systems based on the spatial function defining them. Table 3 illustrates CRIL-nets: The edge marked *IN* represents the function that maps the landmark to its interior region (3.a). *BETWEEN* maps two landmarks to the region that contains all lines connecting the landmarks (3.c) and *LEFT* maps a reference system to the region it identifies as being to the left (3.c; see Eschenbach & Kulik, 1997; Eschenbach, 1999). Paths can be connected to regions via *TO*, *FROM* or *VIA* and to their starting points (*stpt*) and final points (*fpt*) (see section 3.5). The initial CRIL-net is a direct conversion of the propositional specification (see Table 2) to the net-based format.

Table 3. CRIL-representation of example (2): Result of semantic analysis

(2)	(a)	(b)	(c)
Spatial relation	FROM(w_1 , IN(l_1))	TO(w_2 , LEFT($rsys_2$))	VIA(w_3 , BETWEEN(l_2 , l_3))
landmarks	MENSA(l_1)		HOUSE(l_2) NAME(l_2 , 'B') HOUSE(l_3) NAME(l_3 , 'C')
CRIL-net	<pre> graph TD l1["l1 MENSA"] -- IN --> r1["r1"] r1 -- FROM --> w1["w1"] </pre>	<pre> graph TD rsys2 -- LEFT --> r2["r2"] r2 -- TO --> w2["w2"] </pre>	<pre> graph TD l2["l2 HOUSE NAME('B')"] -- BETWEEN --> r3["r3"] l3["l3 HOUSE NAME('C')"] -- BETWEEN --> r3 r3 -- VIA --> w3["w3"] </pre>

Table 3 gives the CRIL-net of the first sentence of example (2) presented in section 2.2. This part of the route instruction describes three paths. The paths are related to regions ((a) *aus der Mensa* [out of the dining hall], (b) *nach links* [to the left] and (c) *zwischen Haus B und Haus C durch* [through [the region] between house B and house C]).

3.5 Knowledge Base: Geometric Concept Specification

Geometric concept specifications (GCS) render geometric aspects of the spatial CRIL-relations precise (Eschenbach et al., 2000). The geometric concept specification includes axiomatic characterizations of spatial concepts, such as *TO*, *FROM* and *VIA* in (D6)–(D8), quoted from Eschenbach et al. (2000). They reveal how a starting point *stpt*(*w*) or a final point *fpt*(*w*) of a path is related to a region, given that the path is related by *TO*, *FROM*, or *VIA* to the region.

$(Q \iota r)$ or $(Q \iota w)$ symbolize that a point (Q) belongs to a region (r) or a path (w), respectively.⁸ The specifications (D6)–(D8) say that a path (w) leads to a region ($\text{TO}(w, r)$) if and only if (iff) the final point of the path belongs to the region ($\text{fpt}(w) \iota r$) but the starting point does not ($\neg(\text{stpt}(w) \iota r)$). A path leads FROM a region iff its starting point belongs to the region but its final point does not. A path leads VIA a region iff both its starting point and its final point do not belong to the region, but another point (Q) of the path does.

(D6) $\text{TO}(w, r) \iff_{\text{def}} \text{fpt}(w) \iota r \wedge \neg(\text{stpt}(w) \iota r)$

(D7) $\text{FROM}(w, r) \iff_{\text{def}} \text{stpt}(w) \iota r \wedge \neg(\text{fpt}(w) \iota r)$

(D8) $\text{VIA}(w, r) \iff_{\text{def}} \exists Q [Q \iota w \wedge Q \iota r] \wedge \neg(\text{stpt}(w) \iota r) \wedge \neg(\text{fpt}(w) \iota r)$

Characterizations in the form of definitions can be used as rewriting rules to transform the CRIL-net. Table 4 shows the result of replacing edges using the definitions (D6)–(D8) as transformation rules. Nodes representing the endpoints of the paths (and one interior point) have been added and related to the regions.

Table 4. CRIL-representation of example (2): Result of access to GCS

(2)	(a)	(b)	(c)
Spatial relation	$\text{FROM}(w_1, \text{IN}(h_1))$	$\text{TO}(w_2, \text{LEFT}(rsys_2))$	$\text{VIA}(w_3, \text{BETWEEN}(h_2, h_3))$
landmarks	$\text{MENSA}(h_1)$		$\text{HOUSE}(h_2) \text{ NAME}(h_2, 'B')$ $\text{HOUSE}(h_3) \text{ NAME}(h_3, 'C')$
CRIL-net			

The geometric concept specification is relevant for different steps during the processing of the CRIL-net. In the instruction phase, the path specifications are used to refine the instruction model. During the navigation phase, for example, the specification of a function as BETWEEN is accessed to determine whether a perceived track is between two perceived houses.

⁸ The Greek letter ι symbolizes the relation of “incidence” basic for incidence geometry (see Eschenbach & Kulik, 1997; Eschenbach et al., 2000).

3.6 Inferences in the Instruction Phase

The Geometric Agent can draw inferences about the route during the instruction phase as well as during the navigation phase. Inferential processing during comprehension, i.e., in advance to navigation, is useful to test one's understanding of the instruction. In spite of that, reasoning involving the real-world constellation of landmarks on the route has to be done during navigation. Nevertheless, the Geometric Agent can serve as a framework to test different distributions of reasoning-load between the two phases.

The succession of actions connects the specifications of the paths in a CRIL-net as displayed above. A useful pragmatic assumption is that the final node of a path is identical to the starting node of the next path to be moved along. Thus, the nodes labeled n_2 and n_3 in the CRIL-net of Table 4 are candidates to be identified. In addition, the specification of w_2 involves an implicit reference system ($rsys_2$). The appropriate choice in this case is to select the directly preceding path w_1 as the crucial direction of $rsys_2$. This results in a CRIL-net as displayed in Figure 3. In a similar way, nodes representing positions (as p_1) can be identified with starting points or final points of paths (in the example n_3). However, the strategies of node identification have to be tested using the Geometric Agent as a model of an instructed navigator.

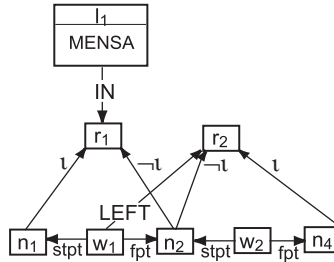


Fig. 3. A CRIL-net resulting from pragmatic processing

Due to the use of projective terms in the instruction, several parameters for reference systems are included in the CRIL-net of example (2). The reference systems $rsys_2$ and $rsys_6$ can be induced by the preceding paths (w_1 and w_4 , respectively). This yields the same result as the explicit instruction to turn left or right, respectively. w_4 is a plausible source for providing $rsys_5$. The intrinsic interpretation of the last projective term corresponds to the identification of the origo of $rsys_7$ and l_4 . All these candidates for reference system parameters can be found during the instruction phase. However, these inferences are based on defaults and therefore they can be withdrawn in case of counter-evidence obtained in the navigation phase.

The interpretation of particles as *wieder [again]* (in example (2)) indicates an implicit change of orientation. In example (2), there are two possibilities to assume an implicit change: either w_3 is right of w_2 or w_4 is right of w_3 . However, the validation of either assumption has to wait until the navigation phase.

Track nodes in CRIL-nets represent perceivable structures in the environment. If the Geometric Agent is restricted to move only on tracks (roads, footpaths), nodes for tracks according to the paths of motion can enrich the CRIL-net. Correspondingly, nodes for paths of motion can extend CRIL-nets generated from descriptions of track constellations.

The Geometric Agent is a formal framework to test methods for constructing internal representations of route instructions and to test the adequacy of these internal representations for navigation. These includes strategies for connecting adjacent paths, for selecting reference systems during the instruction phase, or to enrich the internal representation by implicit actions on explicitly mentioned tracks. Therefore, it can form a systematic approach to pragmatic inferences that enrich the representations resulting from the contribution of lexical semantics.

4 The Geometric Agent in Its Virtual Environment

The Geometric Agent allows studying the interaction between spatial information given in a route instruction and spatial information gained by perception in the course of moving around in the environment. In contrast to mobile robots, which can serve the same purpose, the Geometric Agent idealizes the interaction of an agent with its environment. Object recognition, re-identification of objects perceived at different times, and the detection of object permanence during continuous motion are tasks that cannot be solved in a general way by currently available techniques. The Geometric Agent provides a framework to study instructed navigation independently of such problems of perception.

Tasks of low-level navigation, such as obstacle avoidance or taxis, can be modeled without reference to higher level concepts (Trullier et al., 1997; Mallot, 1999). In the framework of the Geometric Agent, these tasks are part of the simulation of the agent's interaction with the virtual environment. Higher level tasks of navigation addressed in the route instruction have to be mapped to the lower level tasks.

Figure 4 depicts the interaction of the Geometric Agent with the virtual geometric environment. The Geometric Agent's "perceptual model" contains counterpart of objects in the geometric environment (as processed by the agent's perception component) and a plan of the low-level actions. Both the Geometric Agent's perception and its low-level navigation are simulated based on geometric specifications.

The simulation of perception and action bridges the gap between observable spatial behavior and the (propositional) semantics of spatial language. Different components of the agent employ different geometric frameworks. Metric information, for example, is crucial for the simulation of perception and action. Knowledge about distances between objects is also useful to infer directions between the objects when exploring a larger environment (see Trullier et al, 1997). However, route instructions specify directions mostly relative to reference systems and paths relative to landmarks and decision points. Correspondingly, the concepts employed in the CRIL-net that originates from the route instruction belong to affine geometry, whereas in the specification of perception and action metric concepts are employed in addition.

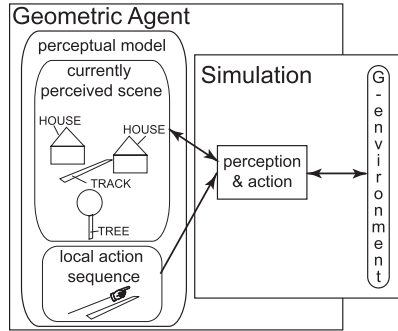


Fig. 4. The interface between the Geometric Agent’s internal representations and its environment

4.1 A Virtual Planar Environment

The geometric model of the environment has two functions. On the one hand, the Geometric Agent perceives parts of the environment and acts in it (see the next section). On the other hand, the environment can be displayed on a computer screen—with the Geometric Agent depicted as small triangle to visualize its current orientation. Thus, a simulation of the Geometric Agent’s actions, i.e., its performing of instructions, can be observed.

The virtual geometric environment of the Geometric Agent is specified in the framework of planar Euclidean geometry. The objects in the virtual environment have geometric properties as shape and pose, represented by points, lines or polygons in the plane. The Geometric Agent is one object in the geometric environment. Its pose is represented by a point⁹ and a half-line (representing its orientation) (Schmidtke et. al, to appear). The geometric properties of the objects are encoded in an absolute coordinate system. In addition, non-geometric attributes as color (GREY), category membership (HOUSE), or label (NAME('B')) specify non-geometric properties the objects.¹⁰

4.2 Simulation of Perception and Action

The perception of the Geometric Agent can be seen as a constructive mapping from the geometric environment into the perceptual model (see Figure 4). The Geometric Agent builds up an internal representation called the “currently perceived scene”. The pose of the Geometric Agent determines a sector of perception. The edges of polygons that intersect this sector and that are not occluded by other edges are recon-

⁹ This idealization of the Geometric Agent as having no extension is suitable since all other objects in the environment can be assumed to be much larger.

¹⁰ Since the virtual environment is planar, the height of objects is represented similar to non-geometric properties.

structed as perceptual objects in the perceptual model. Thus, perceptual objects are the internal representations of the perceivable objects. Depending on spatial parameters, e.g., distance between the Geometric Agent and the objects in the sector of perception, some geometric and non-geometric properties of these objects are transferred to the perceptual model. Similarly, the Geometric Agent can perceive non-geometric properties, such as the name or a salient part of a building, only from certain poses and distances.

The geometrical agent's sector of perception determines which objects of the geometrical environment are perceivable. If the Geometric Agent's perceptual abilities are restricted, then the perceptual model can be imprecise, vague or distorted. Thus, different specifications of the perceptual process can produce different perceptual models.¹¹ The perception module can regard geometric relations corresponding to visual relations (as occlusion) or *gestalt* principles (e.g., objects in a row). The geometric parameters that determine the perceptual mapping, and especially the exactness of this mapping, are controlled and can be changed to test the dependency of the Geometric Agent's performance on these parameters.

The actions of the Geometric Agent and the interdependence of action and perception are controlled in a similar way. In the present stage of modeling, the Geometric Agent is able to approach a distant perceptible target, to follow a perceptible track, and to turn. These abilities correspond to the low-level skills called "taxi", "guidance", and "body alignment" in biological navigation (see Trullier et al., 1997; Mallot, 1999). Since taxi, guidance, and body alignment are low-level navigation skills that are guided by perception, they are simulated based on geometric specifications rather than modeled on the conceptual level.

Higher-level skills of navigation include place recognition, topological or metrical navigation (see Trullier, et al. 1997), and approaching objects that are not perceptible. These skills require that the agent can remember and recognize objects, positions, and constellations of objects. Instructed navigators mainly have to find objects they have never perceived before. Thus, recognition of objects and places described in the instruction is modeled in the Geometric Agent on a higher conceptual level than perception and action.

4.3 The Perceptual Model

The perceptual model contains the currently perceived scene and a current (local) action sequence to be performed. The perceived scene consists of internal representations of perceptible objects, called "perceptual objects". Perceptual objects are representations integrating geometric properties (shape, position) and non-geometric properties like category and color. The Geometric Agent's perception determines which properties of the perceived objects are included in the perceptual model. For example, the perception of the Geometric Agent can directly provide the information that a certain polygon in the geometric environment stands for the region of a house.

¹¹ This separation of perception from the environment corresponds to the two-tiered model presented in Frank (2000).

Objects and parts of objects that are not perceptible from the current pose of the Geometric Agent are not included.

Geometric properties of the perceptual objects are encoded in the perceptual reference system of the Geometric Agent. The relation between the absolute coordinate system of the environment and the Geometric Agent's reference system derives from the Geometric Agent's pose in the environment. Absolute coordinates are not included in the perceptual model. To derive information about its pose relative to objects outside perception, the Geometric Agent has to draw inferences about the environment. The Geometric Agent gathers further information about the environment during the navigation phase and stores it as perceptual enrichment of the nodes in the CRIL-net.

As a second component, the perceptual model contains a projection from the action plan (instruction model), called the "local action sequence". For instance, an instruction like $!GO(w)$ can correspond to a number of low-level navigation actions referring to one or more tracks corresponding to the path w .

5 The Navigation Phase

Instructed navigation requires that two types of internal representations are matched onto each other. The instruction phase yields an internal representation of the route. During the navigation phase, the perception of the environment results in the perceptual model. Both representations contribute to the recognition and identification of an object. The task of recognizing a linguistically described object in the visually perceived scene can—on the theoretical level—be described as the task of co-reference resolution between perceptual objects and nodes in the CRIL-net.

Furthermore, the detection of correspondences between the two internal models enables the Geometric Agent to augment its representation of the environment. This enriched aggregation of instruction model and perceptual model is called the "environment model". It is used for controlling or specifying the action plan (see Figure 5).

The CRIL-net built up from instruction is the initial environment model. In the navigation phase the perceptual model provides new information to enrich the Geometric Agent's internal model of the environment. Due to the augmentation in the navigation phase, the environment model has a hybrid character (Habel, 1987; Habel et al., 1995; Barsalou, 1999). Some parts of the representations are propositional (as $NAME(l_2, 'B')$ or $COLOR(RED)$), other parts contributed from the perceptual model can have an analogous, geometric form.

Spatial reasoning, planning, and high-level navigation are tasks that require knowledge and experience. The environment model provides the spatial information of the Geometric Agent. In the following, we describe three sub-tasks of the navigation phase, which control the processing of the different types of information in the environment model.

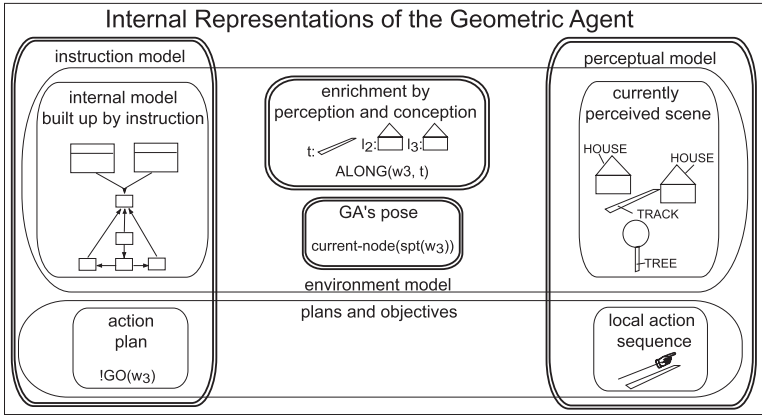


Fig. 5. Internal representations of the Geometric Agent: enriched aggregation of instruction model and perceptual model

Figure 6 depicts the segmentation of the internal representations of the Geometric Agent. The environment model contains the spatial representations gained from perception and route instruction. The level of “plans and objectives” is responsible for triggering the actions of the Geometric Agent. The divisions within the two main segments correspond to the different phases in which the content is acquired.

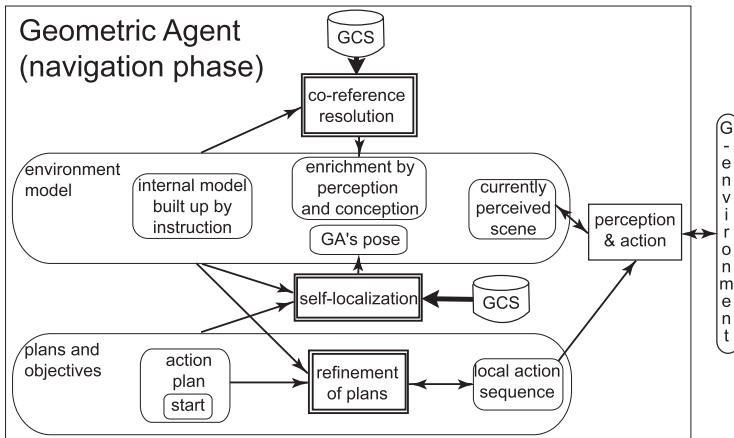


Fig. 6. The internal representations of the Geometric Agent and the central tasks of the navigation phase

Three processes (called “co-reference resolution”, “self-localization”, and “refinement of plans”) access the internal representations. The processes are essentially for carrying out the task of instructed navigation.

Co-reference Resolution. The co-reference resolution has to identify objects from the currently perceived scene with the objects represented in the internal model built up by instruction. This process adds information from the perceptual model and from spatial reasoning to the environment model.

Co-reference resolution bridges the gap between perception and instruction.¹² For example, spatial relations specified in the instruction have to be evaluated in the currently perceived scene to find out which group of perceptual objects fit the description. The CRIL-net and the current pose of the Geometric Agent is evaluated to determine the objects that the Geometric Agent should be able to perceive. The perceptual model is inspected for possible candidates for these landmarks and tracks. The spatial relations specified in the CRIL-net are evaluated in the perceptual model to find the constellation of objects that fit the specification best.¹³ Finally, information from the perceptual model is added to the environment model. This includes further specifications of the recognized objects and additional objects that were not mentioned in the instruction.

Self-localization. Self-localization involves the observation of the progress of the Geometric Agent's pose in the environment model and updating its pose during motion. This is done by comparing expectations about objects and their relations encoded in the CRIL-net with the objects in the currently perceived scene. Self-localization adjusts the estimation of the pose by switching the current node in the CRIL-net whenever a change of pose relative to the environment model has been achieved. A basic requirement of this sub-process is that co-reference resolution has been successful for a sufficient set of perceptual objects. Only after a successful match, the change of geometric relations in the perceptual model can be computed.

Refinements of Plans. The task called the refinement of plans is to supply an action plan that can be carried out by low-level navigation in the geometric environment. Verbal instructions are mostly unspecific in several respects. Therefore, it is necessary to refine the initial action plan and to do local planning. For example, verbal route instructions need not include all decision points but can mention only decision points that require the navigator to turn. Furthermore, the specific shape of lanes or streets to be followed need not be verbally expressed. The refinement of plans component has to ensure a reasonable behavior of the Geometric Agent between the explicitly mentioned decision points. If the Geometric Agent has to follow a longer track, the refinement of plans ensures that the Geometric Agent can act even if the next decision point is not in view.

¹² Frank (2000) describes a system of simulated agents interacting with a simulated environment. It is used to model the complete process of map production and map use. Homomorphisms map between objects of the environment, objects of the map and the corresponding actions of the map-using agent. Since both the instruction model and the currently perceived scene are incomplete, we do not assume co-reference resolution to be a homomorphism.

¹³ Examples for geometric methods to compute the spatial relations for extended objects and objects that are only partially perceived can be found, for example, in Schmidtke (2001).

In the following, we illustrate the function of the three sub-tasks described above with the first sentence of route instruction (2) (compare the representations in Table 2 and Table 3):

Wenn du aus der Mensa kommst, geh nach links, zwischen Haus B und Haus C durch.

When you leave the dining hall, walk to the left, through [the region] between house B and house C.

According to the first phrase, the path w_1 leads from inside the building of the dining hall outside ($FROM(w_1, IN(h_1))$). If the Geometric Agent leaves this building, its trajectory can be identified with w_1 . However, the Geometric Agent needs not to perform this action if it is able to identify the dining hall and a track leading outside. Co-reference resolution has to find such a counterpart of w_1 in the currently perceived scene. In the next step the Geometric Agent has to find a decision point on this track (corresponding to the point where w_1 and w_2 meet), and a track that corresponds to the movement along w_2 . Thus, the co-reference resolution has to determine which track in the perceptual model could be involved in the relation $TO(w_2, LEFT(w_1))$ —given that during the instruction phase $rsys_2$ is identified with w_1 . The process of plan refinement has to introduce the command to move to the decision point and then to align with the track of w_2 . The self-localization process has to specify the pose of the Geometric Agent first relative to w_1 , and later, relative to the point where w_1 and w_2 meet.

The next phrase of the instruction can fit two different spatial constellations. The path w_3 (specified by the relation $VIA(w_3, BETWEEN(l_2, l_3))$) can be a straight continuation of path w_2 , or branch of and lead into another direction. Thus, the paths w_2 and w_3 can correspond to one track in the environment or to two meeting tracks. The co-reference resolution has to map the landmark nodes (l_2, l_3) to perceptual objects and to decide which tracks in the perceptual model fits the description $VIA(w_3, BETWEEN(l_2, l_3))$. The refinement of plans has to form a local plan that reflects the perceived spatial relation between the tracks of w_3 and w_2 . While moving along the tracks, self-localization has to observe and to update the Geometric Agent's pose, for example, it has to give the information needed to decide when the region $BETWEEN(l_2, l_3)$ is entered and left.

6 Conclusion

The Geometric Agent provides a formal framework for investigating higher-level cognitive tasks of instructed navigation. It abstracts from perception and lower-level navigation by simulating both the geometric environment and the agent's interaction with its environment. The paradigm of a Geometric Agent supplies a test-bed for evaluating route instructions and theories on instructed navigation. It allows abstracting from certain details, like a specific architecture and perceptual abilities of a robot.

In the context of in-advance instruction, the processing of the instruction can be carried out during the instruction phase, which is temporally separated from the navigation phase. In the navigation phase, the Geometric Agent has to map the internal propositional representation to the perceived scene. On this basis, it has to identify landmarks and tracks, verify spatial relations among them and find decision points specified in the instruction based on landmarks.

The internal representations proposed in this article derive from investigations on the semantics of natural language, on specific features of route instruction, and on theories on different levels of navigation of animals and robots. The spatial information and the action plan is separated in the representation, but they strongly interact since the commands in the action plan refer to spatial objects such as paths and positions.

The inventory of spatial relations that are underlying the conceptual representation is not fixed. Rather, it is open to additions that are supplied with formal specification of the concepts. Thus, CRIL and the Geometric Agent can be used to evaluate formal studies of the meaning of spatial language in the context of verbal route descriptions. Although the examples discussed in this article focused on natural-language route instructions, CRIL-nets and action plans can be derived from different methods of communicating route instructions, such as sketch-maps or lists of actions. The Geometric Agent provides a framework to compare the different methods of communicating route instructions regarding their adequacy to solve the navigation task.

References

- Agre, P. E. & Chapman, D. (1990). What are plans for? *Robotics and Autonomous Systems*, 6, 17–34.
- Allen, G. L. (1997). From knowledge to words to wayfinding: Issues in the production and comprehension of route directions. In S.C. Hirtle & A.U. Frank (eds.), *Spatial Information Theory* (pp. 363–372). Berlin: Springer.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Crangle, C. & P. Suppes (1994). *Language and Learning for Robots*. Stanford: CSLI.
- Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive* 16, 409–458.
- Eschenbach, C. (1999). Geometric structures of frames of reference and natural language semantics. *Spatial Cognition and Computation* 1, 329–348.
- Eschenbach, C. & L. Kulik (1997). An axiomatic approach to the spatial relations underlying ‘left’–‘right’ and ‘in front of’–‘behind’. In G. Brewka, C. Habel & B. Nebel (eds.), KI-97: Advances in Artificial Intelligence (pp. 207–218). Berlin: Springer-Verlag.
- Eschenbach, C., L. Tschander, C. Habel & L. Kulik (2000). Lexical specifications of paths. In C. Freksa, W. Brauer, C. Habel & K.F. Wender (eds.), *Spatial Cognition II* (pp. 127–144). Berlin: Springer-Verlag.
- Fontaine, S. & M. Denis (1999). The production of route instructions in underground and urban environments. In C. Freksa & D.M. Mark (eds.), *Spatial Information Theory* (pp. 83–94). Berlin: Springer.
- Fraczak, L. (1998). Generating ‘mental maps’ from route descriptions. In P. Olivier & K.-P. Gapp (eds.), *Representation and Processing of Spatial Expressions* (pp. 185–200). Mahwah, NJ: Lawrence Erlbaum.

- Frank, A. (2000). Spatial communication with maps: Defining the correctness of maps using a multi-agent simulation. In C. Freksa, W. Brauer, C. Habel & K.F. Wender (eds.), *Spatial Cognition II* (pp. 80–99). Berlin: Springer-Verlag.
- Habel, C. (1987). Cognitive linguistics: The processing of spatial concepts. *T. A. Informations (Bulletin semestriel de l'ATALA, Association pour le traitement automatique du langage)* 28. 21–56.
- Habel, C., S. Pribbenow & G. Simmons (1995). Partonomies and depictions: A hybrid approach. In J. Glasgow, H. Narayanan & B. Chandrasekaran (eds.): *Diagrammatic Reasoning: Cognitive and Computational Perspectives* (pp. 627–653). Cambridge, MA: MIT-Press.
- Jackendoff, R. (1990). *Semantic Structures*. Cambridge: MIT-Press.
- Kaufmann, I. (1995). *Konzeptuelle Grundlagen semantischer Dekompositionsstrukturen*. Tübingen: Niemeyer.
- Klein, W. (1979). Wegauskünfte. *Zeitschrift für Literaturwissenschaft und Linguistik* 33. 9–57.
- Klein, W. (1982). Local deixis in route directions. In R.J. Jarvella & W. Klein (eds.), *Speech, Place, and Action* (pp. 161–182). Chichester: Wiley.
- Labrou, Y., Finin, T. & Peng, Y. (1999). Agent communication languages: The current landscape. *IEEE Intelligent Systems*, 14. 45–52.
- Levinson, S. (1996). Frames of reference and Molyneux's question: crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett (eds.), *Language and Space* (pp. 109–169). Cambridge MA: MIT Press.
- Ligozat, G. (2000). From language to motion, and back: Generating and using route descriptions. In D.N. Chistodoulakis (ed.) *NLP 2000*, LNCS 1835. pp. 328–345.
- Lovelace, K. L., M. Hegarty & D. R. Montello (1999). Elements of good route directions in familiar and unfamiliar environments. In C. Freksa & D.M. Mark (eds.), *Spatial Information Theory*. (pp. 65–82). Berlin: Springer.
- Mallot, H. A. (1999). Spatial cognition: Behavioral competences, neural mechanisms, and evolutionary scaling. *Kognitionswissenschaft* 8. 40–48.
- Mann, G. (1996). *Control of a Navigating Rational Agent by Natural Language*. PhD Thesis. School of Computer Science and Engineering, University of New South Wales, Sydney.
- Schmidtke, H. R. (2001). The house is north of the river: Relative localization of extended objects. In D. R. Montello (ed.), *Spatial Information Theory* (pp. 414–430). Berlin: Springer.
- Schmidtke, H. R., L. Tschander, C. Eschenbach & C. Habel, (to appear). Change of orientation. In J. Slack & E. van der Zee (eds.), *Representing Direction in Language and Space*. Oxford: Oxford University Press.
- Trullier, O., S. I. Wiener, A. Berthoz & J.-A. Meyer (1997). Biologically based artificial navigation systems: Review and prospects. *Progress in Neurobiology*, 51. 483–544.
- Tversky, B. & P. U. Lee (1999). On pictorial and verbal tools for conveying routes. In C. Freksa & D.M. Mark (eds.), *Spatial Information Theory*. (pp. 51–64). Berlin: Springer.
- Vere, S. & Bickmore, T. (1990) A basic agent. *Computational Intelligence*, 6, 41–61.
- Werner, S., B. Krieg-Brückner & T. Herrmann (2000). Modelling navigational knowledge by route graphs. In C. Freksa, W. Brauer, C. Habel & K.F. Wender (eds.), *Spatial Cognition II* (pp. 295–316). Berlin: Springer-Verlag.
- Wunderlich, D. & R. Reinelt (1982). How to get there from here. In R.J. Jarvella & W. Klein (eds.), *Speech, Place, and Action* (pp. 183–201). Chichester: Wiley.

Cognition Meets Le Corbusier – Cognitive Principles of Architectural Design

Steffen Werner¹ and Paul Long²

¹Department of Psychology, University of Idaho, Moscow, ID, 83844-3043, USA
swerner@uidaho.edu, www.uidaho.edu/~swerner

²Department of Architecture, University of Idaho, Moscow, ID, 83844-2541, USA
long1773@uidaho.edu

Abstract. Research on human spatial memory and navigational ability has recently shown the strong influence of reference systems in spatial memory on the ways spatial information is accessed in navigation and other spatially oriented tasks. One of the main findings can be characterized as a large cognitive cost, both in terms of speed and accuracy that occurs whenever the reference system used to encode spatial information in memory is not aligned with the reference system required by a particular task. In this paper, the role of aligned and misaligned reference systems is discussed in the context of the built environment and modern architecture. The role of architectural design on the perception and mental representation of space by humans is investigated. The navigability and usability of built space is systematically analysed in the light of cognitive theories of spatial and navigational abilities of humans. It is concluded that a building's navigability and related wayfinding issues can benefit from architectural design that takes into account basic results of spatial cognition research.

1 Wayfinding and Architecture

Life takes place in space and humans, like other organisms, have developed adaptive strategies to find their way around their environment. Tasks such as identifying a place or direction, retracing one's path, or navigating a large-scale space, are essential elements to mobile organisms. Most of these spatial abilities have evolved in natural environments over a very long time, using properties present in nature as cues for spatial orientation and wayfinding.

With the rise of complex social structure and culture, humans began to modify their natural environment to better fit their needs. The emergence of primitive dwellings mainly provided shelter, but at the same time allowed builders to create environments whose spatial structure "regulated" the chaotic natural environment. They did this by using basic measurements and geometric relations, such as straight lines, right angles, etc., as the basic elements of design (Le Corbusier, 1931, p. 69ff.) In modern society, most of our lives take place in similar regulated, human-made spatial environments, with paths, tracks, streets, and hallways as the main arteries of human locomotion. Architecture and landscape architecture embody the human effort to structure space in meaningful and useful ways.

Architectural design of space has multiple functions. Architecture is designed to satisfy the different representational, functional, aesthetic, and emotional needs of organizations and the people who live or work in these structures. In this chapter, emphasis lies on a specific functional aspect of architectural design: human wayfinding. Many approaches to improving architecture focus on functional issues, like improved ecological design, the creation of improved workplaces, better climate control, lighting conditions, or social meeting areas. Similarly, when focusing on the mobility of humans, the ease of wayfinding within a building can be seen as an essential function of a building's design (Arthur & Passini, 1992; Passini, 1984).

When focusing on wayfinding issues in buildings, cities, and landscapes, the designed spatial environment can be seen as an important tool in achieving a particular goal, e.g., reaching a destination or finding an exit in case of emergency. This view, if taken to a literal extreme, is summarized by Le Corbusier's (1931) notion of the building as a "machine," mirroring in architecture the engineering ideals of efficiency and functionality found in airplanes and cars. In the narrow sense of wayfinding, a building thus can be considered of good design if it allows easy and error-free navigation. This view is also adopted by Passini (1984), who states that "although the architecture and the spatial configuration of a building generate the wayfinding problems people have to solve, they are also a wayfinding support system in that they contain the information necessary to solve the problem" (p. 110).

Like other problems of engineering, the wayfinding problem in architecture should have one or more solutions that can be evaluated. This view of architecture can be contrasted with the alternative view of architecture as "built philosophy". According to this latter view, architecture, like art, expresses ideas and cultural progress by shaping the spatial structure of the world – a view which gives consideration to the users as part of the philosophical approach but not necessarily from a usability perspective.

Viewing wayfinding within the built environment as a "man-machine-interaction" problem makes clear that good architectural design with respect to navigability needs to take two factors into account. First, the human user comes equipped with particular sensory, perceptual, motoric, and cognitive abilities. Knowledge of these abilities and the limitations of an average user or special user populations thus is a prerequisite for good design. Second, structural, functional, financial, and other design considerations restrict the degrees of freedom architects have in designing usable spaces.

In the following sections, we first focus on basic research on human spatial cognition. Even though not all of it is directly applicable to architectural design and wayfinding, it lays the foundation for more specific analyses in part 3 and 4. In part 3, the emphasis is on a specific research question that recently has attracted some attention: the role of environmental structure (e.g., building and street layout) for the selection of a spatial reference frame. In part 4, implications for architectural design are discussed by means of two real-world examples.

2 The Human User in Wayfinding

2.1 Navigational Strategies

Finding one's way in the environment, reaching a destination, or remembering the location of relevant objects are some of the elementary tasks of human activity. Fortu-

nately, human navigators are well equipped with an array of flexible navigational strategies, which usually enable them to master their spatial environment (Allen, 1999). In addition, human navigation can rely on tools that extend human sensory and mnemonic abilities.

Most spatial or navigational strategies are so common that they do not occur to us when we perform them. Walking down a hallway we hardly realize that the optical and acoustical flows give us rich information about where we are headed and whether we will collide with other objects (Gibson, 1979). Our perception of other objects already includes physical and social models on how they will move and where they will be once we reach the point where paths might cross. Following a path can consist of following a particular visual texture (e.g., asphalt) or feeling a handrail in the dark by touch. At places where multiple continuing paths are possible, we might have learned to associate the scene with a particular action (e.g., turn left; Schölkopf & Mallot, 1995), or we might try to approximate a heading direction by choosing the path that most closely resembles this direction. When in doubt about our path we might ask another person or consult a map. As is evident from this brief (and not exhaustive) description, navigational strategies and activities are rich in diversity and adaptability (for an overview see Golledge, 1999; Werner, Krieg-Brückner, & Herrmann, 2000), some of which are aided by architectural design and signage (see Arthur & Passini, 1992; Passini, 1984).

Despite the large number of different navigational strategies, people still experience problems finding their way or even feel lost momentarily. This feeling of being lost might reflect the lack of a key component of human wayfinding: knowledge about where one is located in an environment – with respect to one's goal, one's starting location, or with respect to the global environment one is in. As Lynch put it, "the terror of being lost comes from the necessity that a mobile organism be oriented in its surroundings" (1960, p. 125.) Some wayfinding strategies, like vector navigation, rely heavily on this information. Other strategies, e.g. piloting or path-following, which are based on purely local information can benefit from even vague locational knowledge as a redundant source of information to validate or question navigational decisions (see Werner et al., 2000, for examples.) Proficient signage in buildings, on the other hand, relies on a different strategy. It relieves a user from keeping track of his or her position in space by indicating the correct navigational choice whenever the choice becomes relevant.

Keeping track of one's position during navigation can be done quite easily if access to global landmarks, reference directions, or coordinates is possible. Unfortunately, the built environment often does not allow for simple navigational strategies based on these types of information. Instead, spatial information has to be integrated across multiple places, paths, turns, and extended periods of time (see Poucet, 1993, for an interesting model of how this can be achieved). In the next section we will describe an essential ingredient of this integration – the mental representation of spatial information in memory.

2.2 Alignment Effects in Spatial Memory

When observing tourists in an unfamiliar environment, one often notices people frantically turning maps to align the noticeable landmarks depicted in the map with the visible landmarks as seen from the viewpoint of the tourist. This type of behavior in-

icates a well-established cognitive principle (Levine, Jankovic, & Palij, 1982). Observers more easily comprehend and use information depicted in “You-are-here” (YAH) maps if the up-down direction of the map coincides with the front-back direction of the observer. In this situation, the natural preference of directional mapping of top to front and bottom to back is used, and left and right in the map stay left and right in the depicted world. While this alignment effect is based on the alignment between the map representation of the environment and the environment itself, alignments of other types of spatial representations have been the focus of considerable work in cognitive psychology. When viewing a path with multiple segments from one viewpoint, as shown in Figure 1, human observers have an easier time retrieving from memory the spatial relations between locations as seen from this viewpoint than from other, misaligned views or headings (Presson & Hazelrigg, 1984). In these types of studies, the orientation of the observer with respect to his or her orientation during the acquisition of spatial information, either imagined or real, seems to be the main factor. Questions like “Imagine you are standing at 4, looking at 3, where is 2?” are easier to answer correctly than “Imagine you are standing at 2, looking at 4, where is 3?”. These results have been taken as an indication of alignment effects between the orientation of an observer during learning and the imagined orientation during test.

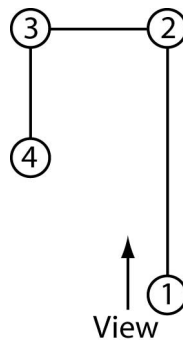


Fig. 1. Sample layout of objects in Presson & Hazelrigg (1984) study. The observer learns the locations of objects from position 1 and is later tested in different conditions.

Later studies have linked the existence of alignment effects to the first view a person has of a spatial layout (Shelton & McNamara, 1997). If an observer learns the location of a number of objects from two different viewpoints he will be fastest and most correct in his response when imagining himself in the same heading as the first view. Imagined headings corresponding to the second view are no better than other, not experienced headings. According to the proposed theory, a person mentally represents the first view of a configuration and integrates new information from other viewpoints into this representation, leaving the original orientation intact. Similar to modern view-based theories of object recognition (Tarr, 1995), this theory proposes that spatial information should be easier accessible if the imagined or actual heading of a person coincides with this “remembered” viewing direction, producing an alignment effect.

In the theories described above, the spatial relation between the observer and the spatial configuration determines the accessibility of spatial knowledge without any

reference to the spatial structure of the environment itself. Indeed, most studies conducted in a laboratory environment try to minimize the potential effects of the external environment, for example by displaying a configuration of simple objects within a round space, lacking in any salient spatial structure. This is in stark contrast to the physical environments a person encounters in real life. Here, salient axes and landmarks are often abundant and are used to remember important spatial information.

Recently, studies of human spatial memory have started to explore the potential effect of spatial structure on human spatial memory and human navigation (Werner, Saade, & Lüer, 1998; Werner & Schmidt, 1999). If an observer has to learn a configuration of eight objects within a square room, for example, she will have a much easier time retrieving the spatial knowledge about the configuration when imagining herself aligned with the room's two main axes parallel to the walls than when imagining herself aligned with the two diagonals of the room. This holds true even when all potential heading directions within the room have been experienced by the observer (Werner, Saade, & Lüer, 1998). Similarly, people seem to be sensitive to the spatial structure of the large-scale environment they live in. When asked to point in the direction of important landmarks of the city they live in, participants have a much easier time imagining themselves aligned with the street grid than misaligned with the street grid (Werner & Schmidt, 1999; see also Montello, 1991). In this case, the environment has been learned over a long period of time and from a large number of different viewpoints. Additional research strongly suggests that the perceived structure of an environment influences the way a space is mentally represented even in cases where the acquisition phase is well-controlled and the observer is limited to only a few views of the space (Shelton & McNamara, 2001; McNamara, Rump, & Werner, *in press*). In sum, the perceived spatial structure of an environment seems to play a crucial role in how spatial information is remembered and how easy it is to retrieve. In the following section we will review which features of the environment might serve as the building blocks of perceived spatial structure.

3 The Perceived Structure of the Environment

Natural and man-made environments offer a large number of features that can influence the perception of "environmental structure." Visual features, such as textures, edges, contours, can serve as the basis for structure as can other modalities, such as sound or smell. Depending on the scale of the environment, the sensory equipment of the user, and the general navigational goal, environments might be perceived very differently. However, in many cases a consensus seems to exist among observers as to the general structure of natural environments. Following are a few examples.

When navigating in the mountains, rivers, valleys, and mountain ranges constitute the dominant physical feature that naturally restrict movement and determine what can be perceived in certain directions. Paths within this type of terrain will usually follow the natural shape of the environment. Directional information will often be given in environmental terms, for example "leaving or entering a valley," "crossing a mountain range," or "uphill" and "downhill" (see Pederson, 1993), reflecting the importance of these physical features. A recent study confirmed that observers use environmental slant not only to communicate spatial relations verbally, but also to structure their spatial memories (Werner, 2001; Werner, Schmidt, & Jajinek, *in prep.*). In

this study, participants had to learn the location of eight objects on a steep hill. Their spatial knowledge of the environment was later tested in the laboratory. Accessing spatial knowledge about this sloped environment was fastest and most accurate when imagining oneself facing uphill or downhill, thus aligning oneself with the steepest gradient of the space.

In many instances, natural boundaries defined through changes in texture or color give rise to the perception of a shaped environment. Looking at a small island from the top of a mountain lets one clearly see the coastal outline of the land. Changes in vegetation similarly present natural boundaries between different regions. Both, humans and other animals seem to be sensitive to the geometrical shape of their environment. Rats, for example, rely heavily on geometrical structure when trying to retrieve food in an ambiguous situation (Cheng & Gallistel, 1984; Gallistel, 1990). Young children and other primates also seem to favor basic geometrical properties of an environment when trying to locate a hidden toy or buried food (Hermer & Spelke, 1994; Gouteux, Thinus-Blanc, & Vauclair, 2001). The importance of geometric relations might be due to the stability of this information over time, compared to other visual features whose appearance can change dramatically throughout the seasons (bloom, changing and falling of leaves, snow cover; see Hermer & Spelke, 1996).

Different species have developed many highly specialized strategies to structure their environment consistently. For migrating birds, local features of the environment are as important as geo-magnetic and celestial reference points. Pigeons often rely on acoustical or olfactory gradients to find their home (Wiltschko & Wiltschko, 1999). The desert ant *Cataglyphis* uses a compass of polarized sunlight to sense an absolute reference direction in its environment (Wehner, Michel, & Antonsen, 1996). Similarly, humans can use statistically stable sources of information to create structure. When navigating in the desert, the wind direction or position of celestial bodies at night might be the main reference, whereas currents might signal a reference direction to the polynesian navigator (see Lynch, 1960, pp. 123ff, for anecdotal references).

In the built environment, structure is achieved in different ways. At the level of the city, main streets and paths give a clear sense of direction and determine the ease with which spatial relations between different places or regions can be understood (Lynch, 1960). In his analysis of the “image of the city,” Lynch points out the difficulty to relate different parts of Boston because the main paths do not follow straight lines and are not parallel. The case of Boston also nicely illustrates the interplay between the built and natural environment. In Boston, the main paths for traffic run parallel to the Charles river – resulting in an alignment of built and natural environment. As mentioned above, the perceived structure of the city plays a large role in how accessible spatial knowledge is for different imagined or real headings within the space (Werner & Schmidt, 1999). At a smaller scale, individual buildings or structures impose their own structure. As Le Corbusier notes, “architecture is based on axes” which need to be arranged and made salient by the architect (p. 187). Through these axes, defined by walls, corridors, lighting, and the arrangement of other architectural design elements, the architect communicates a spatial structure to the users of a building. Good architectural design thus enables the observer to extract relevant spatial information. This feature has been termed architectural legibility and is the key concept in research on wayfinding within the built environment (Passini, 1984, p. 110). In the last section we will focus on the issue of architectural legibility and how the design of a floor plan can aide or disrupt successful wayfinding.

4 Designing for Navigation

4.1 Architectural Legibility and Floor Plan Complexity

Research linking architectural design and ease of navigation has mainly focused on two separate dimensions: the complexity of the architectural space, especially the floor plan layout, and the use of signage and other differentiation of places within a building as navigational aids. As many different research projects have shown both from an architectural and environmental psychology point of view, the complexity of the floor plan has a significant influence on the ease with which users can navigate within a building (O'Neill, 1991, Weisman, 1981, Passini, 1984).

The concept of complexity, however, is only vaguely defined and comprises a number of different components. Most often, users' ratings of the figural complexity of a floor plan, often interpreted as a geometric entity, has been used to quantify floor plan complexity for later use in regression models to predict navigability. Different authors have mentioned different underlying factors that influence an observer's judgment of complexity; most notably, the symmetry of a plan and the number of possible connections between different parts of the figure. An attempt to quantify the complexity of a floor plan analytically, by computing the mean number of potential paths from any decision point within the floor plan, was devised by O'Neill (1991).

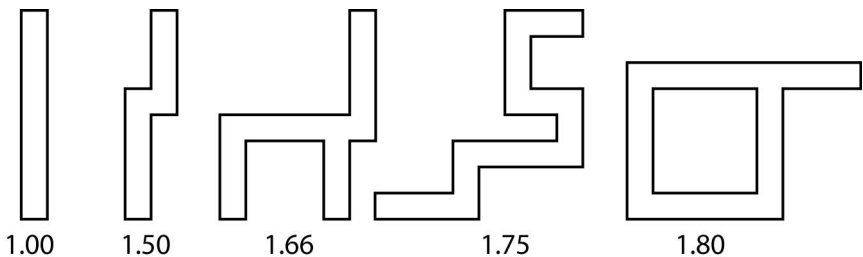


Fig. 2. Different schematic floor plans and their ICD index after O'Neill (1991).

Five basic floor plan layouts used in his study are shown in Figure 2 and the corresponding inter-connection density index (ICD) is listed underneath each plan. The basic idea in this approach consists of an increase in floor plan complexity with increasing number of navigational options or different paths. The correlation of the ICD measure and empirical ratings of complexity for the plans used in his study were fairly high. One theoretical problem with this index, however, is demonstrated in Figure 3. Here 4 different figures depict three different floor plans with exactly the same ICD index. Their perceived complexity, however, rises from left to right, by making the figures less symmetric, changing the orientation, or making the figure less regular.

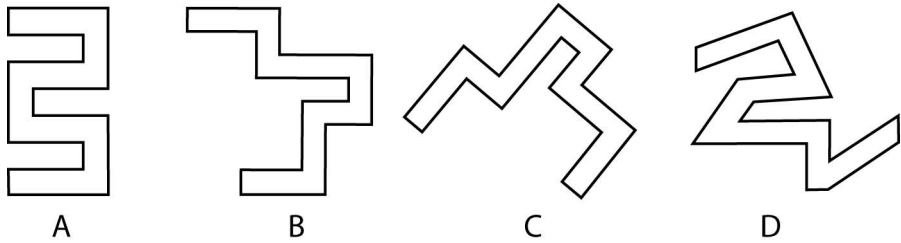


Fig. 3. Four different floor plans with identical ICD but different perceived complexity.

A serious problem with all approaches using figural complexity as a measure, is to treat the geometrical complexity of a floor plan as indicative of the navigational complexity of the spatial environment depicted by the plan. As Le Corbusier pointed out almost 80 years ago, the easily perceivable and pleasant geometrical two-dimensional depiction of a spatial environment can differ dramatically from the perceived structure of a spatial environment (1931, p. 187). In it, space is experienced piecemeal, from multiple different viewpoints, in which only small portions of the space are visible at one time, and in which spatial relations have to be inferred by integrating spatial knowledge across multiple viewpoints and over long periods of time. The basic city layout of Karlsruhe, for example, includes as its main design characteristic a radial (star) arrangement of streets emanating from the castle in the center of the environment. While providing a very salient structure when looking at the city map, the global structure is hidden from each individual view. What is perceived is often a single, isolated street. In a similar fashion, when judging the complexity of the two fictitious floor plans at the top of Figure 4, The left floor plan might be judged as less complex than the right floor plan. This is due to the meaningfulness of the left geometrical figure. If a person has to navigate this floor plan without prior knowledge of this structure, however, the meaningfulness will not be apparent, and the two floor plans will be perceived as similar in their navigational complexity (see the two views from viewpoints within the two floor plans in the lower half of Figure 4). These examples strongly suggest that the two-dimensional, figural complexity of a depiction of a floor plan should not uncritically be taken as a valid representation of the navigational complexity of the represented spatial environment.

4.2 Global and Local Reference Frames in Perceiving Spatial Layout

When viewing a visual figure, such as a depiction of a floor plan, on a piece of paper or a monitor, the figure can usually be seen in its entirety. This allows an observer of the floor plan to see the spatial relations between different parts of the plan, which cannot be perceived simultaneously in the real environment. One of the first steps in the interpretation of the visual form consists of the assignment of a common frame of reference to relate different parts of the figure to the whole (Rock, 1979). There are multiple, sometimes competing solutions to the problem of which reference frame to assign to a figure. For example, the axis of symmetry might provide a strong basis to select and anchor a reference frame in some symmetric figures, whereas the viewpoint of the observer might be chosen for a less symmetric figure. In general, the dis-

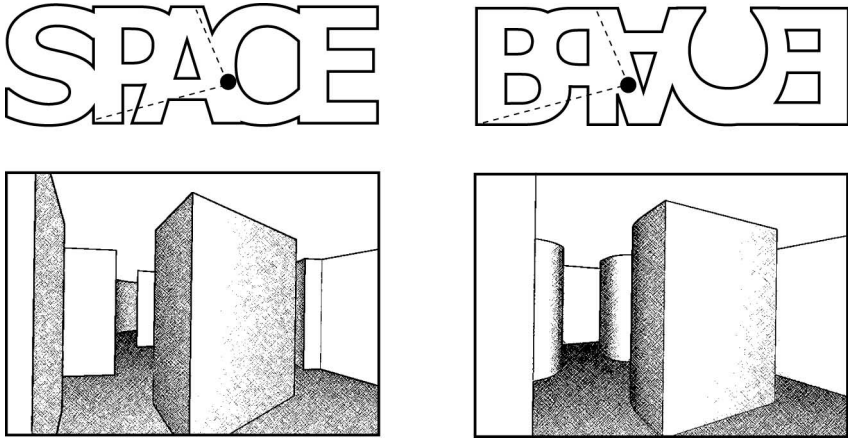


Fig. 5. Two similar floor plans with different perceived complexity; Below: Views from similar viewpoints within the two floor plans (viewpoints and viewing angles indicated above).

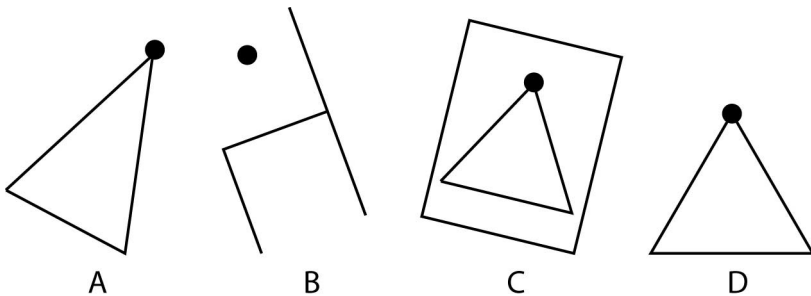


Fig. 5. Determining the “top” of a geometrical figure. Figures A & B exemplify the role of intrinsic reference systems and C & D the role of extrinsic reference systems. The perceived orientation of each figure is marked with a black circle. See text for details.

inction between intrinsic and extrinsic reference frames has proven useful to distinguish two different classes of reference systems.

Intrinsic Reference Systems. An intrinsic reference system is based on a salient feature of the figure itself. In Figure 5 a number of examples illustrate this point. The axis of symmetry in a isosceles triangle determines the perceived direction the triangle is pointing at (example A). It also determines how spatial information within the triangle and surrounding space is organized (e.g., left half and right half, see Schmidt & Werner, 2000). Example B shows a situation in which the meaning of the object determines a system of reference directions (e.g., above and below the chair, see Carlson, 1999). An isolated experience of a particular part of a building will most likely result in the dominance of the intrinsic reference system of the particular space.

Extrinsic Reference System. Besides intrinsic features of a figure, the spatial and visual context of a figure can also serve as the source for a reference system. In example C, the equilateral triangle is seen as pointing towards the right because the rectangular frame around it strongly suggests an orthogonal reference system and only one of the three axes of symmetry of the triangle is parallel to these axes. Similarly, example D shows how the perceived vertical in the visual field or the borders of the page are used to select the reference direction up-down as the most salient axis within the rightmost equilateral triangle. When viewing a floorplan, all the parts of the building can be viewed in unison and the plan itself can be used as a consistent extrinsic reference system for all the parts.

Based on the distinction between extrinsic and intrinsic reference systems we can now re-examine one of the main differences between a small-scale figural depiction of a floor plan and the large-scale space for navigation which is depicted by it. In the case of the small figure, each part of the figure is perceived within the same, common reference system. This reference system can be based on an extrinsic reference system (e.g., the page the plan is drawn on), or a global intrinsic reference system of the plan (e.g., the axis of symmetry of the plan). The common reference system then determines how each part of the plan is perceived.

4.3 Misalignment of Local Reference Systems as a Wayfinding Problem: Two Examples

In section 2 we discussed navigational strategies and how misalignment with the perceived structure of an environment increases the difficulty for a navigator to keep track of the spatial relations between parts of the environment or objects therein. This concept of misalignment with salient axes of an environment fits very well with the concept of a reference system as discussed above. If an environment's structure is defined by a salient axis, this axis will serve as a reference direction in spatial memory. The reference system used to express spatial relations within this environment will most likely be fixed with respect to this reference direction (see Shelton & McNamara, 2001; Werner & Schmidt, 1999).

As discussed in section 2.2, the task of keeping track of one's location in the built environment often requires the integration of spatial information across multiple places. An efficient way to integrate spatial information consists of the expression of spatial relations within the same reference system (Poucet, 1993). A common reference system enables a navigator to relate spatial information that was acquired separately (e.g., by travelling along a number of path segments). Architectural design can aid this integration process by assuring that the perceived spatial structure in each location of a building suggests the same spatial reference system and is thus consistent with a global structure or frame of reference. This does not imply, however, that buildings have to be organized around a simple orthogonal grid with only right angles. Other, more irregular designs are unproblematic as long as the architect can achieve a common reference system by making common axes salient. The following two examples are illustrating the effects of a common reference system and alignment effects at the scale of an individual building (example 1) and the layout of a city (example 2).

Example 1: The town hall in Göttingen, Germany. Figure 6 depicts a schematic floor plan of the town hall of Göttingen, Germany. Informal interviews with people working in or visiting this building revealed that it is difficult to understand and navigate. The architectural legibility is very low. With respect to the aim of this paper, we will mainly focus on the layout of the floor plan in order to discern how it might impact people's ability to find their way around in the building.

When looking at the floor plan, the building appears to consist of three separate areas. To the left and the right, two large areas stand out. They are almost mirror images of each other and slightly offset against each other. At the top of the floor plan, centered horizontally between these two areas is a smaller, third area which includes the main elevator vertically connecting the floors. This area appears to have a diamond shape in the floor plan. To the left, bottom, and right, this area is connected with the hallways serving the other two main areas. The overall shape of the building appears to consist of two offset octagons merged touching on one side with the diamond shaped elevator area connecting them.

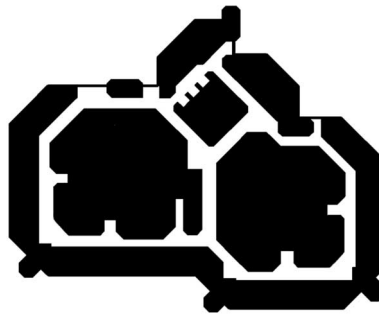


Fig. 6. Floor plan of the city hall of Göttingen, Germany (hallways are depicted in white). The area around the elevator at the top is rotated 45° with respect to the rest of the building.

The naïve description of the visual appearance of the floor plan listed above nicely illustrates the point made above in the context of Figure 4. Especially the description of the elevator area as a “diamond shaped area” needs to be re-evaluated. Unlike a viewer of the floor plan, a user of the physical space will not perceive the area around the elevator as a diamond. Instead, the area will be perceived as a square, thus choosing a different reference system as in the description above. Figure 7 summarizes this situation. Not knowing the global reference system that was used in describing the floor plan, a user upon entering the space will find four hallways surrounding the elevator connected at right angles, leading to the perception of a square.

As is evident from this analysis, an important part of the navigational difficulties in this environment stem from two conflicting spatial reference systems when perceiving different parts of the environment. This misalignment between the parts makes integration of spatial knowledge very difficult.

Example 2: Downtown Minneapolis. The second example deals with a city-scale environment. Figure 8 shows two maps of different parts of downtown Minneapolis. Due to its vicinity to the Mississippi river, the street grid of downtown Minneapolis does not follow the North-South, East-West orientation of the streets and main traffic

arteries found in the surrounding areas. As can be seen in the left map of the warehouse-district, the streets run south-west to north-east or orthogonal to this direction. The map to the right gives an overview of the street grid found downtown and how it connects into the surrounding street pattern (e.g., the streets to the south of downtown).

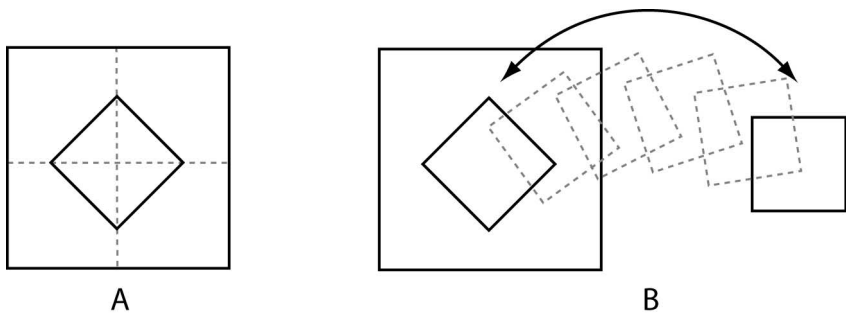


Fig. 7. Schematic display of the spatial situation in the town hall. When viewing image A, the center figure will be labelled diamond. In B, the relation between the figure inside and the outer figure is unknown to the observer and the smaller figure will be seen as a square.

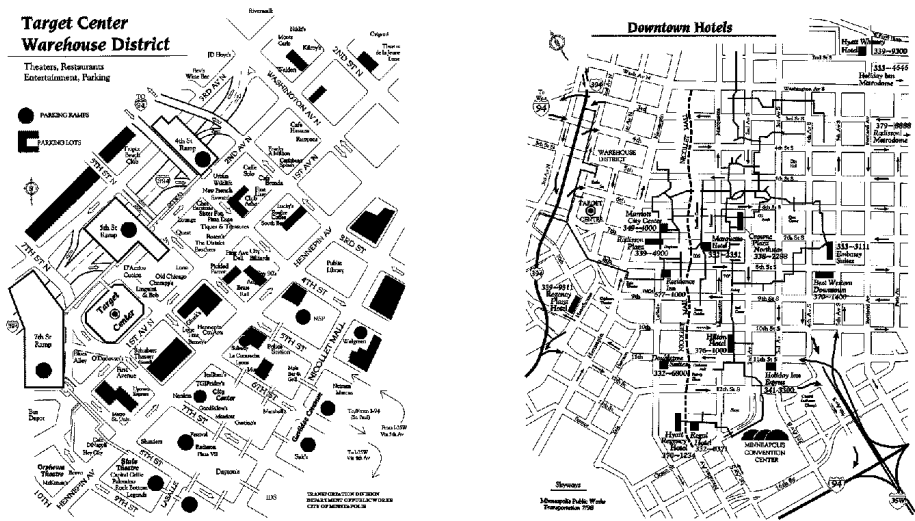


Fig. 8. Maps of downtown Minneapolis. Left: A blown-up map of the Warehouse district. North is up. Note the lack of horizontal and vertical lines. Right: A larger scale depicting all of downtown. In this map, the main street grid consists of vertical and horizontal lines. North is rotated approximately 40° counterclockwise.

It is interesting to note that the map designers for the two maps chose different strategies to convey the spatial layout of the depicted area. On the left, a North-up orientation of the map was chosen, which has the effect that all the depicted streets and buildings are misaligned with the vertical and horizontal. On the right, the map de-

designer chose to align the street grid with the perceived horizontal and vertical on the page, in effect rotating the North orientation by approximately 40° counterclockwise. In a small experiment we tested these types of map arrangements against each other and found that observers had an easier time interpreting and using spatial information gathered from a map in which the depicted information was aligned with the visual vertical and horizontal, whereas a misalignment with these axes led to more errors in judgements about spatial relations made from memory (Werner & Jaeger, 2002). It seems evident, from these results and from the theoretical analysis presented in the context of the town hall, that the information in the map should be presented in the same orientation as it is perceived in the real environment, namely as an orthogonal street grid running up-down, and left-right. The map example on the right also points towards another problem discussed above. When displaying spatial information only about downtown Minneapolis, a rotation of the grid into an upright orientation on the map makes a lot of sense from a usability point of view. However, when this information has to be integrated with spatial information about areas outside the downtown area, the incompatibility of the two reference systems becomes a problem. If information about downtown and the surrounding areas has to be depicted in the same map, only one alignment can be selected (which usually follows the North-up orientation which aligns the streets outside of downtown with the main visual axes).

4.4 Design Recommendations for Wayfinding

As the examples and the discussion of empirical results show, misalignment of reference systems impairs the users ability to integrate spatial information across multiple places. There are a number of design considerations that can be derived from this finding. When designing a building in which wayfinding issues might be relevant, the consistent alignment of reference axes throughout the building, all other things being equal, will greatly reduce the cognitive load while keeping track of once position. The architectural structure as perceived from different locations thus has direct implications for the navigability of the building and determines the buildings overall legibility. Providing navigators access to a global frame of reference within a building will greatly support wayfinding tasks. This can be achieved by providing visual access to distant landmarks or a common link, such as a courtyard or atrium. If the pre-existing architectural environment does not allow for a consistent spatial frame of reference, as in the case of downtown Minneapolis, the navigational demands on the user should take this into consideration. If integration across different reference systems is not required, the problem of misaligned reference systems becomes a moot point. In the case of Minneapolis, for example, the activities in downtown are mainly confined to the regular street grid. Only when leaving the downtown area and trying to connect to the outside street system does the misaligned reference system become an issue. In this case, allowing for simple transitions between the two systems is essential.

Acknowledgements

This paper is based on the results of many empirical studies conducted under a grant to the first author (We 1973/1-3) as part of the priority program on 'Spatial Cognition' funded by the German Science Foundation. The first author wishes to thank all of the students in the spatial cognition lab at Göttingen for their great work. Special thanks go to Melany Jaeger, Vanessa Jainek, Eun-Young Lee, Björn Rump, Christina Saade, Kristine Schmidt, and Thomas Schmidt whose experiments have been mentioned at different parts of the paper. We also wish to thank Andreas Finkelmeyer, Gary Little, Laura Schindler, and Thomas Sneed at the University of Idaho who are currently working on related projects and whose work is also reflected in this paper. Particularly Andreas has been an immense help at all stages of this project.

References

- Allen, G.L. (1999). Spatial abilities, cognitive maps, and wayfinding: Bases for individual differences in spatial cognition and behavior. In R. Golledge (ed.), *Wayfinding behavior* (pp. 46-80). Baltimore: Johns Hopkins.
- Arthur, P. & Passini, R. (1992). *Wayfinding: People, Signs, & Architecture* New York: McGraw-Hill.
- Carlson, L. A. (1999). Selecting a reference frame. *Spatial Cognition and Computation*, 1, 365-379.
- Cheng, K. & Gallistel, R. (1984). Testing the geometric power of an animal's spatial representation. In H.L. Roitblat, T.G. Bever, & H.S. Terrace (Eds.), *Animal cognition* (pp. 409-423). Hillsdale: Erlbaum.
- Gallistel, R. (1990). *The organization of learning*. Cambridge, MA: MIT.
- Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
- Gillner, S. & Mallot, H.A. (1998) Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience*, 10, 445-463.
- Golledge, R.G. (1999). Human wayfinding and cognitive maps. In R. Golledge (Ed.), *Wayfinding behavior* (pp. 5-45). Baltimore: Johns Hopkins.
- Gouteux, S., Thinus-Blanc, C., & Vauclair, J. (2001). Rhesus monkeys use geometric and non-geometric information during a reorientation task. *Journal of Experimental Psychology: General*, 130, 505-519.
- Hermer, L. & Spelke, E. (1994). A geometric process for spatial reorientation in young children. *Nature*, 370, 57-59.
- Hermer, L. & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition*, 61, 195-232.
- Le Corbusier. (1931 / 1986). *Towards a new architecture*. New York: Dover
- Levine, M., Jankovic, I. N., & Palij, M. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology General*, 111, 157-175.
- Lynch, K. (1960). *The Image of the City*. Cambridge: MIT-Press.
- McNamara, T.P., Rump, B., & Werner, S. (in press). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonomic Bulletin & Review*
- Montello, D. R. (1991). Spatial orientation and the angularity of urban routes: A field study. *Environment and Behavior*, 23, 47-69.
- O'Neill, M.J. (1991). Effects of signage and floor plan configuration on wayfinding accuracy. *Environment and Behavior*, 23, 553-574.
- Passini, R. (1984). *Wayfinding in Architecture* New York: Van Nostrand.

- Pederson, E. (1993). Geographic and manipulable space in two Tamil linguistic systems. In A.U. Frank & I. Camari (Eds.), *Spatial information theory* (pp. 294-311). Berlin: Springer.
- Poucet, B. (1993). Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. *Psychological Review*, 100, 163-182.
- Presson, C.C. & Hazelrigg, M.D. (1984). Building spatial representations through primary and secondary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 723-732.
- Rock, I. (1979). *Orientation and form*. New York: Academic Press.
- Schölkopf, B. and Mallot, H. A. (1995). View-based cognitive mapping and planning. *Adaptive Behavior* 3, 311-348.
- Shelton, A.L. & McNamara, T.P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review*, 4, 102-104.
- Shelton, A.L. & McNamara, T.P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology*, 43, 274-310.
- Sholl, M.J. & Nolin, T.L. (1999). Orientation specificity in representations of place. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Sholl, M.J. (1987). Cognitive maps as orienting schemata. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 615-628.
- Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, 2, 55-82.
- Wehner, R., Michel, B., & Antonsen, P. (1996). Visual navigation in insects: Coupling of egocentric and geocentric information. *The Journal of Experimental Biology*, 199, 129-140.
- Weisman, J. (1981). Evaluating architectural legibility: way-finding in the built environment. *Environment and Behavior*, 13, 189-204.
- Werner, S. (2001). Role of environmental reference systems in human spatial memory. *Poster presented at the 42nd Annual Meeting of the Psychonomic Society*, 15-18 November, 2001.
- Werner, S. & Jaeger, M. (2002.). Intrinsic reference systems in map displays. To appear in: *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, Baltimore*.
- Werner, S., Krieg-Brückner, B., & Herrmann, T. (2000). Modelling spatial knowledge by route graphs. In C. Freksa, W. Brauer, C. Habel, & K.F. Wender (Eds.), *Spatial Cognition II - Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications, LNAI 1849* (pp. 295-316). Berlin: Springer.
- Werner, S. & Schmidt, K. (1999). Environmental reference systems for large-scale spaces. *Spatial Cognition and Computation*, 1, 447-473.
- Werner, S. & Schmidt, T. (2000). Investigating spatial reference systems through distortions in visual memory. In C. Freksa, W. Brauer, C. Habel, & K.F. Wender (Eds.), *Spatial Cognition II - Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications, LNAI 1849* (pp. 169-183). Berlin: Springer.
- Werner, S. Schmidt, T., & Jainek, V. (in prep.). The role of environmental slant in human spatial memory.
- Werner, S., Saade, C. & Lüer, G. (1998). Relations between the mental representation of extrapersonal space and spatial behavior. In K.-F. Wender, C. Freksa & C. Habel (Eds.), *Spatial Cognition - An Interdisciplinary Approach to Representing and Processing Spatial Knowledge, LNAI 1404* (pp. 108-127). Berlin: Springer.
- Wiltschko, R. & Wiltschko, W. (1999). Compass orientation as a basic element in avian orientation and navigation. In R. Golledge (ed.), *Wayfinding behavior* (pp. 259-293). Baltimore: Johns Hopkins.

The Effect of Speed Changes on Route Learning in a Desktop Virtual Environment

William S. Albert and Ian M. Thornton

Nissan Cambridge Basic Research, 4 Cambridge Center, Cambridge, MA 02139, USA
William.Albert.1998@alum.bu.edu
Ian.Thornton@tuebingen.mpg.de

Abstract. This study assesses how changes in speed affect the formation of cognitive maps while an observer is learning a route through a desktop virtual environment. Results showed low error rates overall, and essentially no differences in landmark positioning errors between observers in variable speed conditions and a constant speed condition, utilizing both a distance estimation test and mental imagery test. Furthermore, there was a lack of any interactions between speed profiles and trial or route section. These results suggest that the pattern of errors and the nature of learning the route were functionally very similar for both the variable speed conditions and the constant speed condition. We conclude that spatio-temporal representations of a route through a desktop virtual environment can be accurately represented, and are comparable to spatial learning under conditions of constant speed.

1 Introduction

Like many species, humans display great skill in navigating through complex environments. An important part of this skill is the ability to represent aspects of the external world in the form of internal “cognitive” or “mental” maps. The apparent ease with which we construct cognitive maps in the real world is particularly impressive when we consider the variability, both in space (e.g., changes of viewpoint) and time (e.g., changes of speed), which often characterizes our experience within a given environment. The purpose of the current work was to directly assess how changes in speed affect the formation of cognitive maps while an observer is trying to learn a route through a virtual environment.

Changes in speed during navigation are of interest because they modify the relationship between space and time. When speed is held constant during navigation, there is a direct correspondence between the spatial and the temporal separation of landmarks in the environment. When changes in speed occur, however, the two dimensions diverge. For instance, a large distance may be traveled in a short time span or vice versa. Essentially, we know very little about the impact of space - time divergence has on the way we represent the world around.

While it has long been acknowledged that research on cognitive maps should consider the complete “time-space context” of environments (Moore & Golledge, 1976) there has been relatively little empirical work examining how the dimensions of space and time interact during learning. While several studies have examined time

within the context of cognitive mapping (Burnett, 1976; Maceachren, 1980; Säisä, Svensson-Gärting, Gärting, & Lindberg, 1986; McNamara, Halpin, & Hardy, 1992), we know of no other work that has directly assessed the impact that changes in speed might have on both the spatial and the temporal representations of an environment.

The fact that the extensive body of literature on cognitive mapping has paid little attention to "time" is, on the one hand, not very surprising. While the study of time perception itself is well established (see Zakay & Block, 1997 for a recent review), cognitive research in general has typically favored models and metaphors for mental representation, which are inherently static rather than dynamic in nature. Jones (1976) and Freyd (1987) both argued that by omitting a temporal dimension when representing dynamic objects or events (e.g. a musical score, waves breaking on a beach) theories of cognition almost certainly fail to capture fundamental aspects of a world in constant motion and change. The recent growth of interest in connectionism (Rummelhart & McClelland, 1986) and dynamical systems (Berry, Percival, & Weiss, 1987) may help to shift cognitive research away from the idea of purely static representation. As yet, however, temporal aspects of representation are the exception rather than the norm.

On the other hand, the lack of research on temporal aspects of cognitive mapping is surprising when you consider the central role of time in most aspects of real world navigation. For instance, speed generally varies during travel, either as a function of travel mode (e.g., driving, walking, biking, etc.) or environmental conditions (i.e., traffic jams, bad weather, road speed, etc.). Indeed, travel time is often a more significant predictor of spatial behavior than distance (Burnett, 1978). To function effectively in the real world we must constantly compensate for speed changes, taking into account both space and time, in order to develop accurate representations.

The purpose of the current research was to assess the impact that changes in speed might have on an observer's ability to remember the precise location of landmarks within a simple desktop virtual environment. Even desktop virtual reality, in which observers are not fully immersed in an environment, can nevertheless be a useful tool for studying route learning. Observers can be shown exactly the same visual input across multiple presentations, with full control being exercised over the precise playback parameters, such as position on the road or field of view. Furthermore, smooth, continuous motion through the environment can be convincingly simulated and thus the apparent speed of motion, the critical parameter in the current work, can easily be manipulated. In the study described below, we took advantage of this latter point to present separate groups of observers with the same route using different speed profiles. Some observers always experienced the route while traveling at a simulated constant speed. Other groups of observers experienced speed profiles that sped them up or slowed them down during different parts of the route.

Based on previous studies that have used slide presentations rather than a desktop virtual environment (Allen & Kirasic, 1985), we predicted that observers should generally be able to quickly and easily learn the relative position of landmarks within a route. Moreover, their performance should improve with repeated exposure to landmark position. As route learning of this kind involves sequential presentation, we also predicted that the serial position of landmarks within the route would influence performance. That is, items towards the beginning and towards the end of to-be-remembered lists of any kind (e.g., words, pictures, meaningless patterns) usually benefit from what are known as "primacy" and "recency" effects (Postman & Phillips, 1965; Jones, Farrand, Stuart, & Morris, 1995; Avons, 1998). The added saliency of

the endpoints of a list, with less potential for interference from surrounding items, may well account for these effects. In the current context, we might thus expect that memory for the position of landmarks that appear either relatively early or relatively late within the route should be more accurate than memory for items towards the middle of the route.

The main interest in the current study, however, was in whether learning effects or position effects would interact with the speed profile experienced by the observers to influence the precision of landmark placement. Will observers be able to accurately take into account speed changes when making spatial judgments, as would be suggested by real-world performance? Or, will they be biased in their spatial judgments? Examining the errors observers make as they attempt to learn the true *spatial* separation between landmarks as speed is varied, should provide useful insights into how time might affect conceptions of space during navigation.

As mentioned above, changes in speed alter the relationship between the spatial and temporal position of landmarks within a route. One possibility is that this potential “cue conflict” makes it harder for observers to recover the precise location of the landmarks (Rock & Victor, 1964; Steck & Mallot, 2000). For example, observers who experience a variable speed profile will need to adjust their spatial estimates of landmark separation to take into account speed of travel. Specifically, such an observer might be required to expand their spatial judgments during fast speeds, and contract their spatial judgments during slow speeds. Such adjustment could, conceivably, adversely affect performance. On the other hand, observers have a great deal of real-world experience with changes in speed and research from other domains, such as visual-haptic integration suggests that humans can optimally combine cues from different sources (Ernst & Banks, 2002). In this light, we might predict very little deficit for the variable speed conditions, and even possibly some advantage if time and space helped to tune a single representation of the route. Of course, if we find no difference between conditions, this could also reflect a lack of sensitivity in our tests or possible limitations of the current design. We return to this point in the General Discussion.

2 Experiment

Observers were asked to learn the location and time at which they passed by landmarks in a desktop virtual environment. The route was part of a computer-generated environment in which the observer appeared to be traveling as a passenger in a moving vehicle. The route contained a series of nine landmarks in the form of buildings, houses, and other notable structures. Observers were presented the same route six times consecutively. Observers were randomly placed into one of two variable speed conditions, or a constant speed condition. After the first, second and third presentations of the route, observers were simply asked to list the nine landmarks in their correct sequence. Following completion of the fourth, fifth, and sixth trials, all observers completed a distance estimation test and a mental imagery test that required integration of both spatial and temporal knowledge about the route they were presented.

2.1 Observers

A total of 18 observers (6 females and 12 males) participated in the experiment. Observers were compensated for their participation. All observers were naive concerning the purpose of the experiment. All observers were tested in individual sessions lasting one and a half hours.

2.2 Virtual Environment

A virtual environment was created using the GL graphics programming library of a Silicon Graphics Indigo workstation. The virtual environment contained a straight two-lane road with nine landmarks on a textured ground plane. The dimensions of all features and viewing parameters were scaled to be consistent with actual navigation. Therefore, the metrics associated with all features are consistent with their perceived size during actual navigation. The total length of the route was 1,800 meters. Landmarks alternated between the left and right side of the road. The road was 10 meters wide with an intermittent yellow stripe located in the middle of the road to separate the lanes. The landmarks differed from one another in both color and shape, and ranged from 5 meters to 15 meters in height. Landmarks were in the form of office buildings, houses, a monument, fence, and billboard. Landmarks were located approximately 10 meters from the road.

The route was presented as a continuous drive along a straight road, with a viewing height of 1.5 meters above the ground. The maximum viewing distance corresponded to 300 meters. This was controlled by fog that made landmarks beyond 300 meters blend into the background. The route was displayed on a 20" color monitor with a resolution of 1280 (horizontal) x 1024 (vertical) pixels. Subjects sat 65 cm from the monitor, so the display subtended 24 degrees vertically and 41 degrees horizontally. The total duration of this drive was 2 minutes.

2.3 Design

Observers were randomly assigned to one of three speed conditions: SMF (slow-medium-fast), FMS (fast-medium-slow) or MMM (constant medium speed). Therefore, six observers experienced the SMF speed condition, six observers experienced the FMS speed condition, and six observers experienced the MMM (constant speed) condition. The slow speed was equivalent to 10m/second, medium speed was equivalent to 15m/second, and fast speed equivalent to 20m/second. The changes in speed led to a dissociation between the spatial and the temporal placement of landmarks, as shown in Figure 1.

All observers were presented the same route a total of six times. After the fourth, fifth, and six trials, observers participated in two tests: distance estimation and mental imagery. Test order was counterbalanced for all observers. A 3 (speed condition) x 3 (trial) x 3 (route section) experimental design was used. Speed condition was a between-subjects factor containing three groups (SMF, FMS, and MMM). Trial (trial 4, trial 5, and trial 6) and route section (start, middle, and end) were both within-subjects factors.

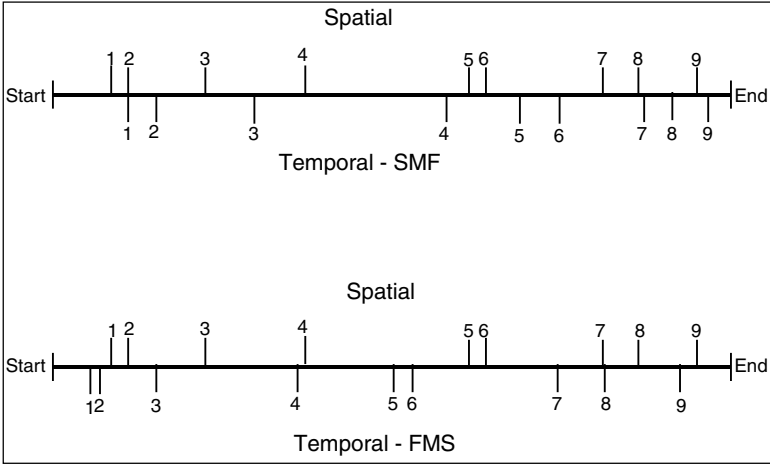


Fig. 1. The upper figure shows the SMF speed profile and the lower figure shows the FMS profile. The top markers on each graph depict the spatial position of landmarks, the bottom markers depict the temporal position of landmarks. As speed of travel is not constant, the spatial and temporal positions do not line up with one another. The MMM speed condition would depict perfect alignment between the spatial and temporal markers (not shown).

2.4 Distance Estimation Test

Observers estimated the relative location of each of the nine landmarks and the two speed changes (with the exception of the constant speed condition). Each measurement was taken by presenting observers with a picture of a landmark and a linear scale from 0 to 100 units. Observers were asked to assume that the total length or distance of the route was 100 units from start to finish. To mark the perceived relative location of each landmark, observers slid the marker along the scale and clicked at the appropriate position. Landmarks were presented in a random order and the initial position of the marker was also randomized before each measurement. Previous measurements remained visible so that the layout of the route was constructed in an incremental fashion. Once all nine landmarks were placed, observers could adjust the position of any of the markers. After the landmarks assignments had been made, the location of the two speed changes was estimated on the same scale using the same method (with the exception of the constant speed condition).

2.5 Mental Imagery Test

In the current work we used a new variant of an imagery task to assess the degree to which spatial and temporal representations of the route can be integrated. Mental imagery has long been used as a tool to probe the nature of mental representation (Podgorny & Shepard, 1978; Kosslyn, 1980; Kosslyn, 1994). Recent work has also begun to use mental imagery as a way to explore the representation of various forms of dynamic events, including navigation through complex environments (Engelkamp

& Cohen, 1991; Ghaem, Mellet, Crivello, Tzourio, Mazoyer, Berthoz, & Denis, 1997; Smyth & Waller, 1998).

Observers were asked to close their eyes and imagine themselves traveling through the route. Each time they “mentally” passed one of the landmarks they pressed the space bar. Observers were told that they could travel at whatever speed felt comfortable, however they should try to take into account the changes in speed. The space bar response was used to estimate the relative locations of each of the landmarks. Accurate performance on such a “mental navigation” task, given that there are changes in speed, requires integration of both the spatial and the temporal dimensions of the learning experience.

2.6 Procedure

The same route was presented a total of six times. Observers were instructed to learn the route as best they could, paying particular attention to the sequence of landmarks, relative location of landmarks, and the speed changes (with the exception of the constant speed condition). After the first three presentations, observers provided a written list of the nine landmarks. This was done in order to verify the correct sequence of landmarks was being learned. Pilot testing indicated that at least three repetitions were necessary to ensure that the speed changes had been noticed and that transposition errors were not made being made in landmark order. These rather conservative checks were necessary to ensure that sufficient learning had taken place for the mental imagery task to provide useful data. After the fourth presentation, of the route, observers participated in each of the two tests. The same tests were also repeated after the fifth and sixth presentations. The purpose of repeating the same two tests after the fourth, fifth and sixth trial was to potentially identify learning effects.

3 Results

Analysis for both distance estimation and mental imagery tests focused on estimates of landmark position. As in previous studies of distance cognition, ratio estimations were computed for both tests. That is, the entire route was normalized to a value of 1.0, with each landmark estimate being placed in its relative position between 0 and 1. For example, a landmark located exactly in the middle of the route would have a value of 0.5. A landmark located close to the end of the route might have a value of 0.9.

Two performance measures were used: absolute error and relative error. Absolute error indicates the magnitude of the difference between the estimated and actual position, without regard to the direction of the difference (over or underestimations). For example, if an observer perceived the location of the centrally located landmark at 0.4, they would have an absolute error of 0.1 or 10%. Relative error is the signed difference between the estimated and the actual landmark location. Given the example above, the observer would have a relative error of -0.1 or -10%. This means that the observer *underestimated* the location of the landmark by 10%. Together, these two measures will indicate the accuracy and the bias of the observer’s estimations.

A 3 (speed condition) x 3 (trial) x 3 (route section)) mixed analysis of variance was performed. The speed condition was a between subjects factor, based on observers being randomly assigned to one of three different speed conditions: SMF, FMS, or MMM. Both trial and route section are within-subjects factors. Trial contains three levels (trial 4, trial 5, and trial 6). Route section also contains three levels (start, middle, and end). Error rates (absolute and relative) were calculated for the first third of the route, middle third of the route, and last third of the route. Each route section contained three landmarks. Main effects and two-way interactions are reported.

3.1 Sensitivity to Speed Changes

Nine of the 12 observers in the variable speed conditions (SMF and FMS) noticed the speed changes on the first trial, with the 3 remaining observers noticing the speed changes by the 2nd or 3rd trials, indicating that observers were aware that their speed was changing. Average absolute error for locating the position of the speed changes by the sixth trial was 8.0% (144 meters) for the first speed change and 9.1% (164 meters) for the second speed change, showing that observers were able to locate these two points to about a hundred and fifty meters. Furthermore, there was very little bias in their estimations of speed change position. Observers underestimated the position of the first speed change by 2.5%, and overestimated the position of the second change by 0.5%.

3.2 Distance Estimation Test

There was no main effect of speed condition for absolute error, suggesting that overall levels of performance were essentially the same with and without a change of speed, $F(2,15) = 0.191$, $MSE = 0.005$, $p = 0.83$. Absolute error rates ranged from 4.3% for the FMS condition to 5.1% for the MMM (constant speed) condition (see Figure 2). For reference, an error of 5% is equivalent to 90 meters.

There was also no main effect of relative error between the three speed groups, $F(2,15) = 0.038$, $MSE = 0.012$, $p = 0.96$. Observers in all three speed conditions slightly underestimated the position of the landmarks, from -2.3% in the MMM condition, to -2.9% in the FMS speed condition (see Figure 2). Such underestimation is not unusual in spatial measures of landmark placement, although it typically occurs when the separation between landmarks is quite large (Holyoak & Mah, 1980)

There was a small, but consistent improvement across trial, $F(2,30) = 4.186$, $MSE = 0.005$, $p = 0.02$. Absolute error rates dropped from 5.5% (99 meters) on trial 4 to 4.0% (72 meters) on trial 6 (see Figure 3). As predicted, observers were able to fine-tune their spatial representation of landmark locations as they became more familiar with the route. A similar pattern of results was also observed for the relative errors. There was a marginally significant reduction in relative error from trial 4 to trial 6, $F(2,30) = 2.956$, $MSE = 0.004$, $p = 0.07$. Observers underestimated landmark position in trial 4 by -3.5%, in trial 5 by -2.9% and trial 6 by -2.4% (see Figure 3). Essentially, observers were beginning to stretch out their representation of landmarks along the route, thus reducing the magnitude of the underestimations.

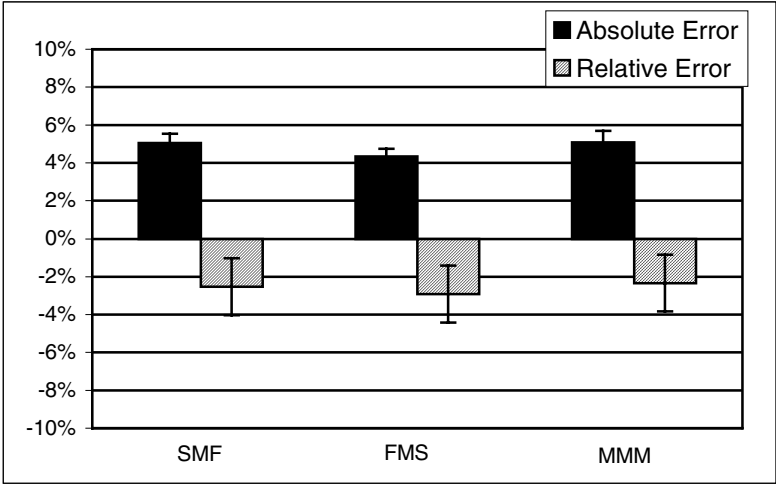


Fig. 2. Average absolute and relative error rates for landmark positioning in the distance estimation test across the three speed conditions. There were no significant differences between the three speed conditions for either absolute or relative errors.

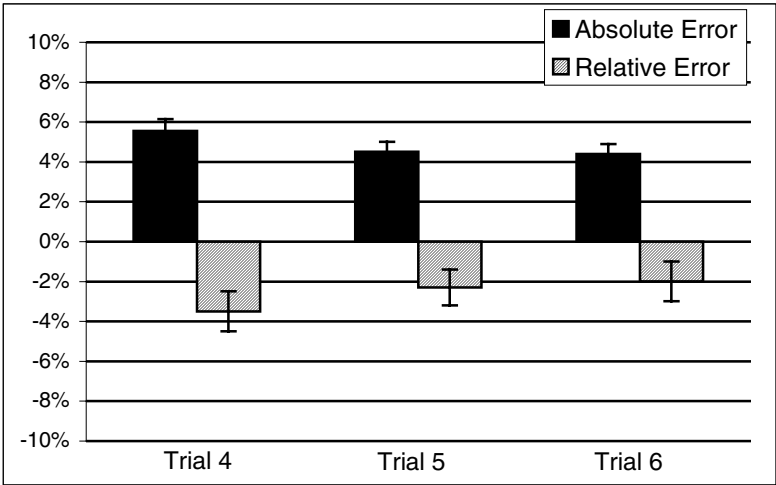


Fig. 3. Average absolute and relative error rates for trial 4, trial 5, and trial 6 in the distance estimation test. There was a significant improvement in performance across trial for absolute error, and a marginally significant improvement for relative error.

There was no significant interaction between trial and speed condition for absolute error, $F(4,30) = 1.863$, $MSE = 0.001$, $p = 0.14$. Observers in all three speed conditions were improving their overall accuracy across trials. However, there was a marginally significant interaction between trial and speed condition for relative error, $F(4,30) = 2.372$, $MSE = 0.002$, $p = 0.08$. Observers in the SMF condition exhibited a relatively greater reduction in bias across trial than either the FMS or MMM conditions.

Observers in the SMF condition reduced the bias in their estimations from -4.4% to -1.2%. However, observers in the FMS condition actually did not reduce their bias in their estimations (-2.3% in trial 4 to -2.6% in trial 6). Observers in the MMM condition had only a slight reduction in the bias of their estimations, from -3.7% on trial 4 to -2.1% on trial 6.

The route was broken into three sections: The start (landmarks 1-3), middle (landmarks 4-6), and the end (landmarks 7-9). The start and end sections were experienced at different speeds by all three conditions. The middle section was experienced in the medium speed by all three conditions. An examination of the three route sections showed a significant main effect for absolute error, $F(2,60) = 9.673$, $MSE = 0.004$, $p < 0.001$. Performance was most accurate in the first third (start) of the route (2.9%), and least accurate on the last third of the route (6.6%). Performance was also more accurate on the last third of the route (5.0%), than the middle third of the route (see Figure 4). This pattern of results is consistent with the serial position effects discussed in the Introduction, indicating a strong primacy effect, and, to a lesser degree, a recency effect.

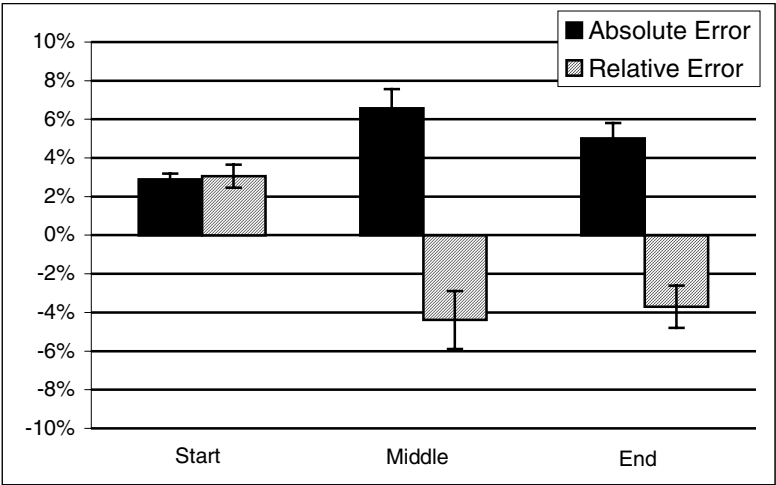


Fig. 4. Average absolute and relative error rates for the start, middle, and end of the route in the distance estimation test. Absolute error rates at the start of the route were significantly better than the middle or end of the route.

A significant main effect was also observed for relative error, $F(2,60) = 8.686$, $MSE = 0.004$, $p < 0.001$. Observers were over-estimating the landmarks in the first route segment by +3%, underestimating the landmarks by -4.4% in the middle third of the route, and also underestimating landmark locations by -3.7% in the last third of the route. It is unclear why the distances at the beginning of the route should tend to be overestimated. Perhaps the combination of the relatively small physical separation between the initial pair of landmarks (see Figure 1; Holyoak & Mah, 1980) and the additional saliency of the start of each trial contribute to this pattern. For example, initial onsets of events are known to attract attention (Yantis & Jonides, 1980) and the allocation of attention has been shown to alter the subjective experience of time (Tse,

Intrilligator, Cavanagh, & Rivest, 1980). Together, these factors could have affected the subjective experience of distances, leading to expansion at the beginning of the route.

There was no significant interaction between landmark position and speed condition for absolute error, $F(4,30) = 1.596$, $MSE = 0.001$, $p = 0.20$. Observers in the three speed conditions were all most accurate in the first route section, and least accurate in the middle route section. Also, there was no significant interaction between landmark position and speed condition for relative errors, $F(4,30) = 0.497$, $MSE = 0.004$, $p = 0.74$. Observers were biased in the same general manner in the three route sections, despite their different temporal experiences with the route.

In summary, the current distance estimation test was unable to show any significant difference in performance between the three speed groups. While this may reflect on the general efficiency with which space and time can be integrated, we cannot rule out the possibility that the current method of testing was simply not sensitive enough. Performance was close to ceiling in all conditions, a factor that could be masking potential differences. Having said this, we were able to demonstrate clear learning effects across trial, suggesting that there was some potential for performance differences. Nevertheless, it is possible that the trend for a learning $\times$ speed profile interaction would have reached significance with a little more statistical power. More generally, while the distance estimation test is useful for measuring the observer's spatial representation of the route, it does not directly measure their temporal representation of the route. The mental imagery test may therefore prove to be a more sensitive test since it requires the observer to actively integrate both their spatial and temporal representations of the route.

3.3 Mental Imagery Test

There was a marginal significant main effect of speed condition on absolute error rates in the mental imagery test, $F(2,15) = 3.017$, $MSE = 0.001$, $p = 0.08$. This marginal main effect reflects relatively poor performance in the SMF speed condition (see Figure 5). Performance ranged from 3.4% in the MMM speed condition up to 9.0% in the SMF speed condition. In addition, there was a significant main effect of speed condition for relative errors, $F(2,15) = 4.024$, $MSE = .0019$, $p = 0.04$. The observers in the SMF condition underestimated the position of landmarks by 8.3%, while the observers in the FMS and MMM speed conditions underestimated the landmarks by 1.5% and 2%, respectively.

This general pattern of underestimation is consistent with the pattern observed in the distance estimation task, and thus could reflect an essentially spatial error. On the other hand, given the nature of the task, this pattern of relative errors could also reflect a temporal or spatio-temporal error. In the general human timing literature, short durations, as used here, tend to be fairly accurately reproduced, whereas longer intervals do tend to be underestimated (Eisler, 1980; Zakay & Block, 1997). Underestimation is also common with other forms of temporal tasks, such as time to collision, where underestimation of imagined spatial-temporal intervals increases greatly with interval size (Schiff & Detwiler, 1979).

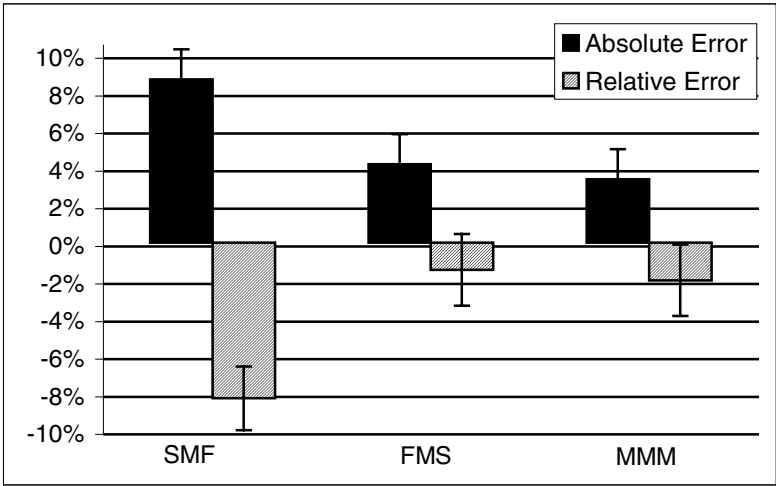


Fig. 5. Average absolute and relative error rates for the three speed conditions in the mental imagery test. The SMF speed condition is significantly worse than either the FMS or constant speed (MMM) conditions for relative error, and marginally worse for absolute error.

It is unclear as to why observers in the SMF performed poorer in the mental imagery test than observers in the other two speed conditions. It is possible that mental “acceleration” is more demanding task than maintaining a constant velocity or mentally “decelerating”. An answer to this question must await future research. In any event, the lack of a difference between the FMS and MMM speed conditions suggests that speed changes per se do not negatively impact the ability to develop an accurate spatio-temporal representation of the route. That is, observers in the FMS condition were able to successfully integrate both spatial and temporal information during their mental reproduction of the route.

There was a marginal main effect of trial on absolute error rates, $F(2,15) = 2.86$, $MSE = .002$, $p = .07$. Similar to the distance estimation test, absolute errors rates were higher in trial 4, however there was little change between trial 5 and trial 6 (see Figure 6).

Observers did not gain any extra accuracy in their landmark estimations from trial 6. It is interesting to note the absolute error rates for trial 5 and trial 6 are only slightly higher (.045 in the distance estimation test and .05 in the mental imagery test). This suggests that both the distance estimation test and mental imagery test are relying on the integration of spatial and temporal representations of the route. Unlike the distance estimation test, relative error did not improve across trial, $F(2,15) = 0.714$, $MSE = 0.001$, $p = 0.50$. Observers underestimated the position of landmarks to the same degree, even as they became more familiar with the route. Perhaps the act of actively integrating both spatial and temporal information during the mental imagery task produces greater distortions of the route.

Similar to the distance estimation test, there was a significant main effect of route section on absolute error rates, $F(2,15) = 17.14$, $MSE = .0015$, $p < 0.001$. Absolute error rates were significantly worse in the middle section of the route, and best for the start of the route. Absolute error rates were slightly under 4% for the start of the

route, and about 5% for the end of the route, and 7.5% for the middle section (see Figure 7). An examination of the route section for the relative errors produced a similar pattern of results, $F(2,15) = 13.687$, $MSE = 0.003$, $p < 0.001$. Observers showed the smallest amount of bias in the first section (-1.2%), and the largest amount of bias on the middle section of the route (-5.8%). This finding shows that changes in speed or a constant speed does not impact the primacy and recency effects using either the distance estimation test or mental imagery test.

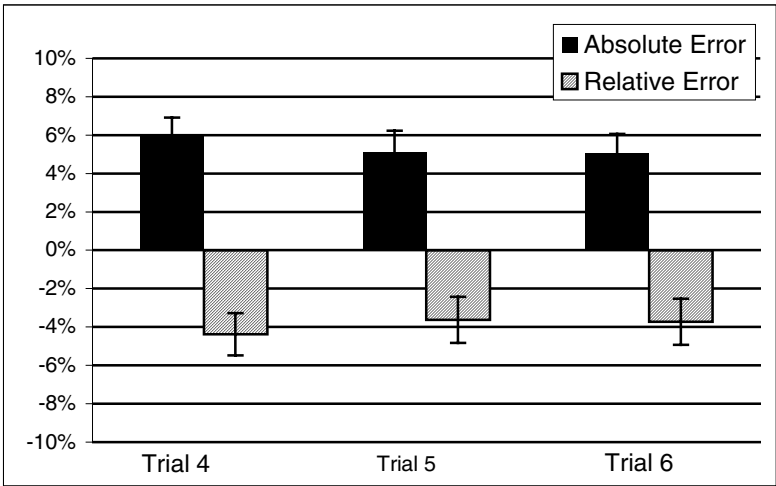


Fig. 6. Average absolute and relative error rates for across trial in the mental imagery test. There was a marginally significant improvement in absolute error rates across trial. However, there was not any improvement in relative error rates across trial.

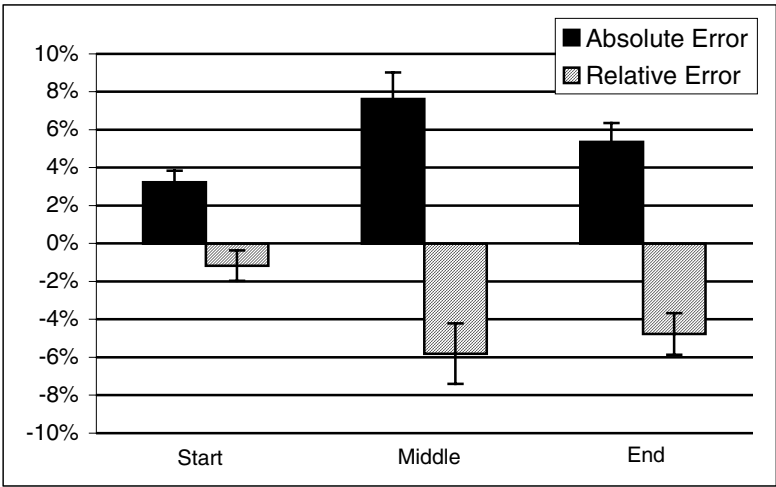


Fig. 7. Average absolute and relative error rates for the three route sections in the mental imagery test. Absolute error rates on the middle section of the route were significantly worse than the start or end sections.

There was a lack of interaction between route section and speed condition for absolute error, $F(4,30) = 1.862$, $MSE = 0.002$, $p = 0.14$. All three speed conditions showed the smallest amount of absolute error in the first section of the route, and the largest amount of absolute error in the middle section of the route. An examination of the relative errors showed a slightly different pattern. There was a marginally significant interaction between route section and speed condition for relative error, $F(4,30) = 2.655$, $MSE = 0.002$, $p = 0.052$. Observers in the FMS condition actually had the largest amount of underestimation on the end section of the route (-3.4%) compared to the middle section (-1.5%). Observers in both the SMF and MMM conditions were least biased in the beginning section of the route (-4.0% and -0.9%, respectively) and most biased in the middle section of the route (-12.4% and -3.6%, respectively).

4 General Discussion

The purpose of this study was to investigate the impact that changes in speed might have on an observers' ability to learn the relative positions of landmarks within a virtual route. In general, all observers, regardless of speed profile, were able to perform very accurately in both a standard distance estimation test and a novel form of mental imagery task. Error rates never exceeded 10% and all observers showed clear performance improvements with repeated exposure. We suggest that this generally high level of performance reflects the frequent exposure and relative ease with which a spatial and temporal experience can be integrated.

Nevertheless, at least in the current environment, we were able to detect subtle differences between traveling at constant versus variable speeds. Specifically, observers in our SMF variable speed group performed significantly worse on the mental imagery task and showed a trend towards a different pattern of learning in the distance estimation test. Interestingly, the second variable speed group, FMS, produced levels of performance that were comparable with, if not a little better than the constant speed group. This indicates that speed variability per se does not necessarily degrade performance and hints at more subtle interactions between the particular spatial and temporal parameters of a route. Consistent with this notion, while the absolute level of performance of the FMS group remained high in the imagery task, the pattern of relative errors across the different sections of the route differed from the SMF and MMM groups. Together, these results suggest that while temporal variation may not strongly bias spatial estimates, and vice versa, the integration of these two sources of route information is not cost free, and certainly does not lead to performance advantages, at least in the current environment.

Clearly, the current study is only a first step in exploring the interaction between time and space during route learning. While some differences between the speed groups have been identified, the current design does not allow us to precisely determine why particular combinations of route position and speed modulate performance. Furthermore, the near ceiling levels of performance – possibly due to the simplicity of our route or the multiple testing sessions -- raise the possibility that we are underestimating the impact of the spatial and temporal dissociation brought about by changes in speed. Also, we clearly cannot rule out the possibility that under some circumstances, perhaps under high spatial uncertainty, variable speed conditions

could afford a performance advantage. In general, however, we speculate that differences between constant and variable speed conditions will remain subtle and relatively hard to detect, even with design improvements that maximize sensitivity. Indeed, it may only be with tests such as the imagery task introduced here -- tests that are sensitive to both spatial and temporal parameters -- that any form of difference will be detectable.

In fact, we believe that the main contribution of the current study is the introduction of dynamic mental imagery task as a tool for studying cognitive mapping. This task, by its very nature, forces a connection between the spatial and the temporal experience of a route. If observers perform this task as instructed, they will re-experience the spatial layout of the route using a specific temporal pattern. While it is not possible to guarantee that observers are re-experiencing the remembered spatio-temporal layout, as opposed to performing a simple time reproduction task, neuroimaging studies (Ghaem et al, 1997) suggest that brain areas involved in the perception of visual scenes become active during such imagery tasks. Also, during debriefing, all observers reported that they had been attempting to mentally navigate through the route. Performance on the imagery task was generally very good, particularly in the constant speed condition, which produced the lowest error rates observed across all tasks and conditions. This suggests some form of coherent representation linking the spatial and temporal aspects of the route. Further studies will be needed to establish whether this performance is supported by a distinct spatio-temporal memory of the route (Freyd, 1987) or whether it reflects the efficient on-line integration of separate spatial and temporal representations. One way to examine this issue would be to test the flexibility of the underlying representation(s). For instance, observers could be asked to "mentally navigate" the route using a different speed profile to the one experienced during learning (e.g., SMF or FMS observers could be asked to imagine traveling at a constant speed) or to traverse the route starting at different points or in reverse order.

There are a number of other ways in which the current work could be usefully extended. First, it would be interesting to apply the empirical methods developed here to learning in a more naturalistic setting, in which people physically navigate through an environment while closely controlling their speed of travel. Second, either in a virtual or a real environment, we could manipulate the amount of control observers have over their exploration of the environment. In the current task observers are always passive. Would active exploration further enhance performance? For instance, would speed changes be more or less salient if they were under the control of the observer? The issue of active versus passive navigation is beginning to be explored in VR environments (Cutmore, Hine, Maberly, Langford, & Hawgood, 2000), but as yet, we know of no studies that show major modulation in performance. Third, our current task repeats the same speed profile on each trial. A more rigorous test of the impact of speed on spatial estimates would be to use a unique speed profile during each exposure. If accurate spatial estimates could still be obtained under these conditions, this would provide even stronger evidence that accurate spatial representations can evolve in the context of temporal variation.

Finally, both of the testing methods employed in the current study involve estimates of time or position with reference to the entire route. A more rigorous method of testing might involve isolated landmark-to-landmark estimations. Such testing could also explore the flexibility of the underlying representations. For instance, how would estimates change if the judgment required a reversal in direction

or time or required judgments across intervening landmarks? Such tests would also shed light on whether the observed tendency to underestimate landmark position is context free or context sensitive.

In conclusion, we believe the current work makes several important contributions to cognitive mapping, both in terms of the empirical approach we have taken and in our attempts to focus attention on the temporal as well as the spatial dimension of navigation. Previous studies of cognitive mapping have generally not varied speed or have not controlled for speed of motion as an experimental factor. Thus, this work represents, to our knowledge, the first direct test of cognitive mapping across changes in speed. Second, our inclusion of explicit tests of time as well as space, brings the field closer to the goal of exploring the complete "time-space context" of environments (Moore & Golledge, 1976). Our main finding, that changes of speed have only subtle impacts on our ability to represent space or time in a virtual world, appears to be very consistent with intuitions gained from everyday navigation.

References

- Allen, G. W. and Kirasic, K.C. (1985). Effects of the cognitive organization of route knowledge on judgments of macrospatial distance. *Memory and Cognition*, 13, 218-227
- Avons, S. E. (1998). Serial report and item recognition of novel visual patterns. *British Journal of Psychology*, 89, 285-308
- Berry, M., Percival, I., & Weiss, N. (Eds.) (1987). *Dynamical Chaos*. NJ: Princeton University Press
- Burnett, P. (1978). Time cognition and urban travel behavior. *Geografiska Annaler*, 60B, 107-115
- Cutmore, TRH, Hine, T. J., Maberly, K. J., Langford, N. M, Hawgood, G. (2000). Cognitive and gender factors influencing navigation in a virtual environment. *International Journal of Human-Computer Studies*, 53, 223-249
- Eisler, H. (1980). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, 83, 1154-1171
- Engelkamp, J., & Cohen, R. L. (1991). Current issues in memory research. *Psychological Research*, 53, 175-182
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429-433
- Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94, 427-438
- Ghaem O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A., & Denis, M. (1997). Mental Navigation along memorized routes activates the hippocampus, precuneus and insula. *NeuroReport*, 8, 739-744
- Holyoak, K. J., & Mah, W. A. (1980). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, 14, 328-352
- Jones, D. M., Farrand, P., Stuart, G. P., & Morris, N. (1965). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 1-11
- Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological Review*, 83, 323-355
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press
- MacEachren, A. M. (1980). Travel time as the basis of cognitive distance. *Professional Geographer*, 32, 30-36

- McNamara, T. P., Halpin, J. A., & Hardy, J. K. (1992). Spatial and temporal contributions to the structure of spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 555-564
- Moore, G. T., & Golledge, R. G. (1976). *Environmental Knowing*. Stroudsburg, Pennsylvania: Dowden, Hutchinson & Ross
- Podgorny, P., & Shepard, R. (1978). Functional representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 21-35
- Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 17, 132-138
- Rock, I., & Victor, J. (1964). Vision & touch: An Experimentally created conflict between the two senses. *Science*, 143, 594-596
- Rummelhart, D. E., McClelland, J. L., and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol 1: Foundations. MA: MIT Press
- Säisä, J., Svensson-Gärling, A., Gärling, T., & Lindberg, E.(1986). Intraurban cognitive distance: The relationship between judgments of straight-line distances, travel distances, and travel times. *Geographical Analysis*, 18, 167-174
- Schiff, W. & Detwiler, M.L. (1979). Information used in judging impending collision. *Perception*, 8, 647-658
- Smyth, M. M., & Waller, A. (1998). Movement imagery in rock climbing: Patterns of interference from visual, spatial and kinaesthetic secondary tasks. *Applied Cognitive Psychology*, 12, 145-157
- Steck S. D., & Mallot H. A. (2000). The role of global and local landmarks in virtual environment navigation. *Presence-Teleoperators and Virtual Environments*, 9, 69-83
- Tse, P., Intrilligator, J., Cavanagh, P., & Rivest, J. (1980). Attention distorts the perception of time. *Investigative Ophthalmology & Visual Science*, 38, S1151
- Yantis, S., & Jonides, J. (1980). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception & Performance*, 10, 601-621
- Zakay, D., & Block, R. A. (1997). Temporal Cognition. *Current Directions in Psychological Science*, 6(1), 12-16

Is It Possible to Learn and Transfer Spatial Information from Virtual to Real Worlds?*

Doris Höll¹, Bernd Leplow², Robby Schönfeld², and Maximilian Mehdorn¹

¹Clinic for Neurosurgery, Christian-Albrechts-University of Kiel, Weimarer Str. 8,
24106 Kiel, Germany

dhoell@psychologie.uni-kiel.de

²Department of Psychology, Martin-Luther-University of Halle, Brandbergweg 23,
06099 Halle (Saale), Germany

b.leplow@psych.uni-halle.de

Abstract. In the present study spatial behavior was assessed by utilization of a desktop virtual environment and a locomotor maze task. In the first phase of the experiment, two groups of healthy middle-aged participants had to learn and remember five out of 20 target locations either in a “real” locomotor maze or an equivalent VR-version of this maze. The group with the VR-training was also confronted with the task in the real maze after achieving a learning criterion. Though acquisition rates were widely equivalent in the VR- and locomotor groups, VR participants had more problems learning the maze in the very first learning trials. Good transfer was achieved from the virtual to the real version of the maze by this group and they were significantly better in the acquisition phase of the locomotor task than the group that had not received VR-training. In the second phase of the experiment -the probe trials- when the cue configuration was changed the group with the VR-training seemed to have specific problems. A considerable number of participants of this group were not able to transfer information.

Key Words: spatial cognition, spatial orientation, spatial memory, memory, orientation, VR-environment

1 Introduction

In the last few years various studies have shown the potential of virtual reality (VR) technology not only to train technical staff, but also for clinical purposes (Reiss & Weghorst, 1995; Rose, Attree, & Johnson, 1996; Rizzo & Buckwalter, 1997; Antunano & Brown, 1999). VR allows us to see, to hear, and to feel a world created graphically in three dimensions, and to interact with it. This world could be imaginary or inaccessible for us and VR also allows us to construct environments in which we can completely control all the stimuli and alter them to the needs of the person experiencing this world. The user is not only an observer of what is happening on a screen, but he immerses himself in that world and participates in it, in spite of the fact

* This research was supported by the DFG governmental program “Spatial Cognition” (Le 846/2-3).

that these spaces and objects only exist in the memory of the computer and in the user's mind (immersion is a term that refers to the degree in which a virtual environment submerges the user's perceptive system in virtual stimuli). It is designed to simulate diverse effects directed to one or sometimes even more senses, with the purpose that the virtual world will come closer to the real world.

There are some examples of real life situations that come to mind when thinking of using this technology. Imagine for example emergency training for a ship's crew. It is possible to train their ability to orientate themselves in VR under extremely difficult and dangerous conditions without putting the crewmember in real danger. A fire for example could badly impair vision or a vessel which has a tilt of perhaps 20° could lead to extreme problems in finding the way to the upper decks and the life-boats. Simulating these conditions in the real world would be quite expensive and extremely complicated.

There are also examples in the field of clinical psychology and neurological rehabilitation where the use of VR has been tested (Johnson, Rose, Rushton, Pentland, & Attree, 1998). It is possible to provide a wide range of motor responses in everyday situations in people whose motor disability restricts their movement in the real world. Examples of this are people using wheelchairs or patients with Parkinson's disease, who have severe deficits with their movements and therefore have to use economical strategies. For these patients VR-training could help to learn about new environments without wasting too much energy. Another example of the use of VR is given by Emmett (1994) who used the knowledge that, despite their difficulty in walking, Parkinson's patients do indeed step over objects placed in their paths, so by superimposing virtual obstacles on the real environment normal gait was achieved.

Another group of patients that could benefit from this training means are patients who suffer from spatial memory and orientation deficits. Standard neuropsychological orientation tests are very often paper pencil tests and could be too narrow and artificial to give an accurate impression of these cognitive functions in real life situations. In order to get a more ecologically valid measure, in a virtual environment we could create a realistic situation and have the opportunity to maintain strict control over every aspect of this test situation. Also we can create an atmosphere in which we directly observe a patient's behavior and what is happening to this person. The interaction between the user and the environment gives us the advantage of a participant who is more than a mere observer but an actor on his own stage. Skelton, Bukach, Laurance, Thomas, and Jacobs (2000) used computer-generated environments on patients with traumatic brain injuries who showed place-learning deficits in a computer-generated virtual space. Performances in the virtual environment correlated with self-reported frequency of way finding problems in everyday life and with scores on a test of episodic memory of the Rivermead Behavioral Memory Test (Wilson, Cockburn, & Baddeley, 1985). Certainly VR has the potential to improve on existing assessments of sensory responsiveness, to maximize the chance of identifying the right combination of stimuli and to minimize the chance of missing a meaningful response.

Another advantage is the ability to use neuroimaging of spatial orientation tasks in a computer-generated virtual environment. Thomas, Hsu, Laurance, Nadel, and Jacobs (2001) could demonstrate that all training procedures effectively taught the participants the layout of a virtual environment and also the application of a computer-generated arena procedure in neuroimaging and neuropsychological

investigation of human spatial navigation. But still the question of what we actually measure arises. How similar are the cognitive processes of spatial orientation and memory if a maze task is computer generated or performed in a real world environment?

Also some other problems have to be kept in mind. First of all vision and sound are the primary feedback channels in most of the studies today that have been published. Other setups provide more sensory information such as tactile information in data gloves or body suits but these technologies however are quite expensive at present and require a considerable amount of further development and research in order to use them with patients. Another problem that has been reported in many studies is a form of motion sickness that has been termed "cybersickness" or simulator sickness. Cybersickness is believed to occur when there is a conflict between perception in different sense modalities (auditory, visual, vestibular, or proprioceptive) or when sensory cue information in the VR environment is incongruent with what is felt by the body or with what is expected based on the user's history of real world sensory experience. In a study by Regan and Price (1994), 61% of 146 healthy participants reported "symptoms of malaise" at some point during a 20-minute immersion and 10-minute postimmersion period causing 5% of the participants to withdraw from the experiment before completing their 20-min immersion period. And this side effects issue is of particular importance when considering the use of VR for persons with neurological injuries, some of whom display residual equilibrium, balance, and orientation difficulties.

And a question that still remains to be answered is to what degree a sense of immersion has to be created in the participants' senses in order to have a useful tool e.g. for the training of virtual environments or assessment of spatial orientation and memory. Riva (1998) distinguishes between immersive VR and virtual environments (VE) saying that VR is characterized by an immersive technology using head-mounted displays, and interaction devices, such as data-gloves or a joystick. Whereas VE may be displayed on a desktop monitor, a wide field of view display such as a projection screen. A VE is fixed in a space and is referred to as partially immersive by this author. Input devices for these desktop environments are largely mouse and joystick based.

In the study we report about in this paper the computer generated virtual environment is presented solely on a conventional computer monitor and the participant navigates through the world by means of a joystick. This does not create such a big sense of immersion as multi wall stereo projection systems or a Head-Mounted-Display (Mehlitz, Kleinoeder, Weniger, & Rienhoff, 1998). In this field of research it is common opinion today that desktop-systems are as effective as immersive systems in some cognitive tasks. A weighty reason to use this mode of presentation is to reduce the rate of participants experiencing cybersickness reported in other studies. An additionally important reason to use desktop-VR is that it is our future objective to use this technology on patients in hospitals and clinics and therefore it needs to be mobile.

Another question that also arises when using VR technologies is, whether spatial information that was acquired in a VR environment can be transferred into real life situations. Are there any problems that could appear because of missing proprioception and vestibular input? And how complex or simple can a VR environment be in order to provide enough visual information for the participant to transfer this information into real life? In their study Foreman, Stirk, Pohl,

Mandelkow, Lehnung, Herzog, and Leplow (2000) addressed the questions whether spatial information that is acquired in a virtual maze transfers to a real version of the same maze. Foreman and colleagues used a VR version of the Kiel locomotor maze, a small scale space which is described in detail in a preceding volume of this book (Leplow, Höll, Zeng, & Mehdorn, 1998). In the study of Foreman et al. enhanced acquisition of the task in the real world was observed in 11-year-old children following an accurate training in an early virtual version of the same maze. The virtual version of the maze was presented on a desktop computer monitor. The authors found that good transfer was achieved from the virtual to the real version of this maze. Children made fewer errors and learned more rapidly than children without any training and even children, who received misleading training before entering the real maze were able to transfer information into the real world and performed better than the group of children that did not receive any training in advance. The authors conclude that it is clear that transfer of spatial information occurs from the simulated Kiel maze to the real version. Rose and colleagues (1997) support this view in their study. They found that positive transfer can occur between virtual and real environments when using a simple sensorimotor virtual task in VR. When transferred to a real world task the participants benefited from virtual training as much as from real practice. This paper addresses the question of details regarding transfer of information that was acquired in a VR environment into a real world environment in adults. Do learning rates obtained from participants who learned in real world environments differ from those who acquired the spatial layout in virtual space? Is VR-based spatial training sufficient if environmental configurations within the real world have changed considerably?

2 Method

2.1 Apparatus

Real Life Environment (Kiel Locomotor Maze). Participants were asked to explore a room of 4.16 x 4.16 m. It was the participants' task to identify and remember five out of 20 hidden locations on the floor of this room. These locations were distributed on the floor in a semi-irregularly fashion and were marked by very small light points, inserted into the floor next to a capacity detector (Fig. 1a). This detector can register the presence of a human limb and therefore can record the track of spatial behavior in this chamber. The detectors were connected individually to a microcomputer in the neighboring room that automatically registered each participant's behavior.

The light points could only be seen when a subject positioned himself about 30 cm away from that detector and therefore only about 2 to 3 light points could be seen at a time which prevented the participants from using geometric encoding strategies. This arrangement was used following the 'hidden platform' paradigm by Morris (1981).

The whole room was painted black and soundproof so that participants were prevented from orienting themselves according to acoustic stimuli from outside the experimental chamber. In each of the corners of the chamber extramaze cues with clearly distinguishable abstract symbols of about 30 x 30 cm in size were provided. In order to have cues that provide the same visual information during the whole length

of the experiment we replaced the fluorescent symbols used in earlier studies (Lepow et al., 1998, 2000) with circuit boards that were equipped with light-emitting diodes. No other cues were visible except for two proximal cues of about 5x5x5 cm that were also fitted with symbols made of light-emitting diodes. These cues were provided in order to investigate different navigational strategies in the probe trails and are located at predefined positions on the floor of the experimental chamber.

It was the participants' task to step on the locations. The five locations that were defined as "correct" locations emitted a 160 Hertz-tone when activated by the participants' limb. A second step on one of these correct locations did not yield another tone. An error was recorded. When activating the other 15 incorrect locations by stepping on them no feedback tone was provided and an error was also recorded. After two subsequent errorless trials the acquisition phase was completed.

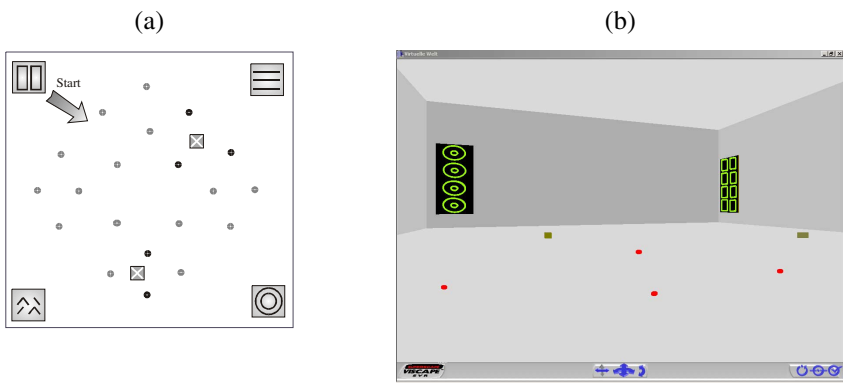


Fig. 1. Spatial layout of (a) locomotor maze and (b) VR-environment

VR-Environment. The VR-environment was equivalent to the locomotor maze described above (Fig. 1b). Only this time the chamber to be explored was displayed on a standard 17 inch computer monitor and the participant could move around in this chamber with a joystick. The same cues were provided and it was the participants' task to explore the room and find and remember five out of 20 locations. This was done by crossing the locations with an area on the bottom of the screen that was defined as the participant's foot. If the participant found a correct location a pleasant chime-like sound was elicited. In case of an error an unpleasant humming sound could be heard. This was made to ensure the participants that the location was successfully visited. Again it was the participants' task to make two consecutive errorless trials in order to finish an acquisition phase.

2.2 General Procedure

After informed consent had been obtained, participants were guided to the locomotor maze or the computer version of the maze was opened. The participants of the first group were guided into the locomotor maze and given the instructions to explore the room, visit each location, step onto each detector and to try and remember the correct locations (those locations that elicit a feedback tone). After the first exploration trial

the participants were asked to try and visit the correct locations only. The acquisition phase was over when the participants successfully finished two consecutive trials without making any error.

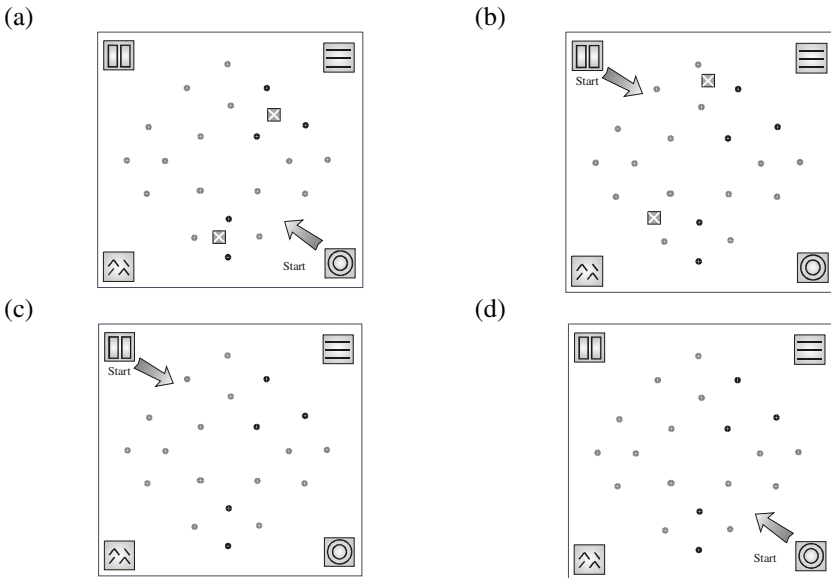


Fig. 2. Probe trials (a) Test 1: “response rotation”, (b) Test 2: “cue rotation”, (c) Test 3 “cue deletion”, (d) Delay: “response rotation and cue deletion”

Then the participants were blindfolded and disorientated and guided to a new starting position within the chamber, where the first test trial was started (“response rotation”, Fig. 2a). Again it was the task to find the five correct locations. For the second test (“cue rotation”, Fig. 2b) the participant again was disorientated as described above and led to the starting position of the learning phase. While blindfolded the proximal cues were rotated by 180° and the second test was started. After the participant had found the correct locations again she or he was disorientated and this time the proximal cues were removed (“cue deletion”, Fig. 2c). The subject was led to the same starting position and had to find the five correct detectors again. The last test (“delay”, Fig. 2d) was performed after an interval of about 30 min. The participant again was led to the starting position and had to find the five correct locations in order to finish this task.

The second group started the experiment with the VR-maze (Fig. 3). The computer monitor was placed in a dimly lit room in order to reduce distractions by light reflexes on the screen or the surrounding furniture and other things that were placed in that room. Before entering the VR-maze the participant had to enter a so called “waiting room”. In this room the use of the joystick was trained to have equal starting conditions for each of the participants. When she/he felt comfortable with the handling of the joystick and had finished a simple motor task in the waiting room, the VR-maze was opened. The maze was an exact copy of the locomotor maze described above. Again it was the participants’ task to find and remember five out of 20 hidden

locations. The acquisition phase was over after two consecutive trials without errors. After finishing the acquisition phase in the VR-maze the participants were led into the locomotor maze. Again they had to try and find the five correct locations without making any error. After finishing the acquisition phase in the locomotor maze successfully participants then were also exposed to the tests described above. In both the VR- and the locomotor maze, the acquisition phase was terminated if the participant had spent more than 30 minutes in either version of the maze.

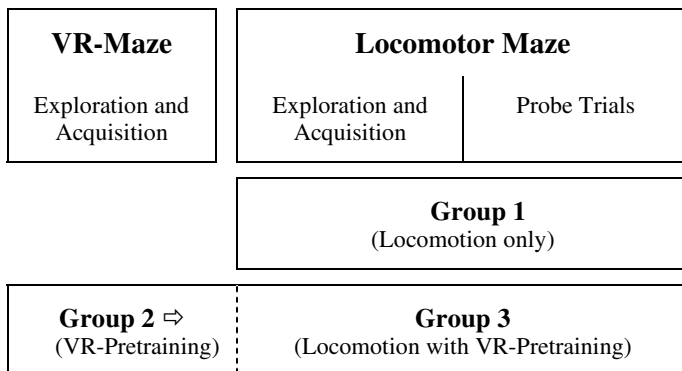


Fig. 3. Design of the Experiment

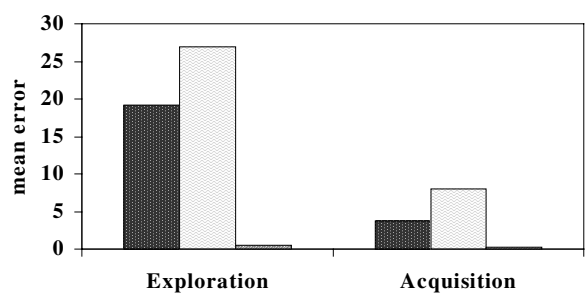
2.3 Participants

Two groups of middle-aged right-handed healthy participants closely matched for age and intelligence were recruited. A level of intelligence was estimated using the MWT-B (Lehrl, 1975), a test of word recognition which is functionally equivalent to the widely used NART test (Nelson & O'Connell, 1978). One group of 16 participants (eight females) did the locomotor maze task only. The mean age in this group was 45.75 years. 16 participants (eight females) of the second group reached the learning criterion in the VR-task. One subject (6.25%) who participated in the VR-training reported feelings of slight nausea after finishing the VR-task, and therefore did not complete the transfer task into the locomotor maze. The mean age of the participants who performed the transfer task was 43.38.

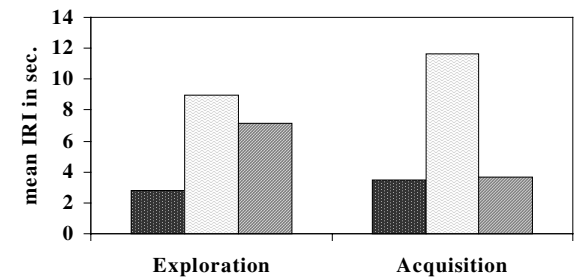
3 Results

No sex differences were observed in either group in the variables trials to learning criterion, spatial memory errors and inter-response intervals (IRI = mean time elapsing between visiting each location within one trial). Therefore both sexes were combined for further analysis.

(a)



(b)



■ Locomotion only □ VR-Pretraining ▨ Locomotion with VR-Pretraining

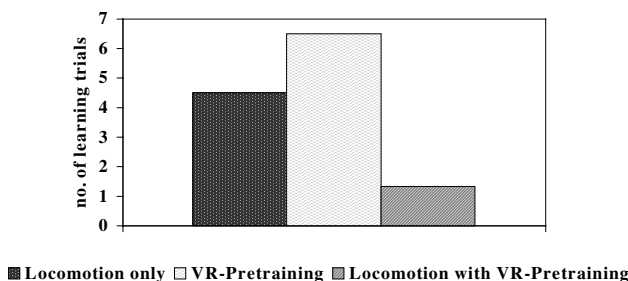
Fig. 4. Mean errors (a) and mean IRIs (b) within exploration and acquisition phase

Exploration Behavior.

For the exploration phase the average number of false locations visited was examined between three groups (Fig. 4a). Group 1 consisted of participants who went into the real maze only (locomotion only), group 2 consisted of participants who received a VR-training (VR-training) and group 3 consisted of participants who went into the real maze after receiving the VR-training (locomotion with VR-pretraining). Therefore group 2 and 3 actually consisted of the same participants (Fig. 3). Group 1 visited 19.25 locations, group 2 visited 27 detectors and group 3 stepped onto 0.50 locations.

Analysis showed that group 1 visited more locations than group 3 ($z = -4.96$, $p = 0.00$), but not than group 2 ($z = -1.55$, $p = 0.12$). Participants in group 2 also stepped on significantly more locations than group 3 ($z = -4.96$, $p = 0.00$). In the inter-response intervals (IRI, Fig. 4b) group 1 needed a mean IRI of 2.82 seconds, group 2 yielded an IRI of 8.96 seconds and group 3 needed an average of 7.18 seconds for two subsequent visits. Group 1 was significantly faster than group 2 ($z = -4.71$, $p = 0.00$) and group 3 ($z = -4.10$, $p = 0.00$), whereas no differences could be observed between group 2 and 3 ($z = -1.39$, $p = 0.17$).

(a)



(b)

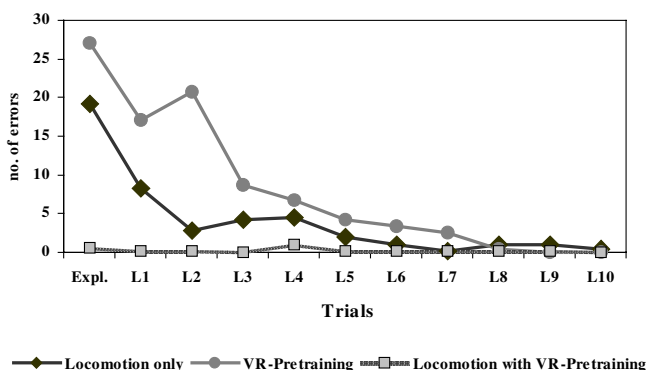


Fig. 5. The mean number of learning trials (a) and the course of errors during (b) acquisition (across exploration and learning trials)

Acquisition Rates. In this phase of the experiment participants who were confronted with the locomotor maze only made an average of 3.75 errors, participants who received a VR-training collided with 7.96 locations on average and participants who went into the locomotor maze after a training phase in VR performed an average of 0.25 errors in acquisition. Analysis showed that while there were no significant differences observed between groups 1 and 2 ($z = -1.70$, $p = 0.088$) group 1 and 3 ($z = -3.26$, $p = 0.001$) and groups 2 and 3 ($z = -4.23$, $p = 0.000$) differed significantly in this measure (Fig. 4a). When having a closer look at the course of errors during acquisition (Fig. 5b) in the first two trials significant differences could be observed between group 1 and 2 (trial 1: $z = -1.98$, $p = 0.047$; trial 2: $z = -3.06$, $p = 0.002$). This could be an indication that although on average no difference of errors in the acquisition phase was observed the participants who were confronted with the VR-maze required a more extensive search in early acquisition behavior in order to master the task.

In the IRI measure participants of group 2 showed the longest interval with 11.62 seconds on average (Fig. 4b) and were significantly slower than group 1 with 3.46 sec. ($z = -4.52$, $p = 0.000$) and group 3 with 3.67 sec. ($z = -4.48$, $p = 0.000$). Groups 1 and 3 did not differ ($z = -0.65$, $p = 0.57$). The number of trials to reach the learning

criterion of two consecutive errorless trials (this measure includes these trials) differed significantly between group 1 and group 2 ($z = -2.11$, $p = 0.035$). Participants of group 1 needed fewer trials than group 2 to reach the criterion (Fig. 5a). Participants of group 2 also needed significantly more acquisition trials than group 3 ($z = -4.67$, $p = 0.000$) in this measure. Analysis showed that group 1 also needed significantly more learning trials than group 3 ($z = -4.31$, $p = 0.000$) to end the acquisition phase. In the real maze participants mastered the task in an average of 4.5 trials, in the VR-maze the mean number of trials was 6.5 and in the transfer task it took the participants an average of 1.33 trials to reach the learning criterion.

Probe Trials. Only group 1 and 3 participated in the probe trials. In the first probe trial (response rotation) the participants with a learning experience only in the locomotor maze performed an average of 0.56 errors with an IRI of 2.95 seconds (Fig. 6). The group that had had an acquisition phase in the VR-maze and the locomotor maze thereafter made an average of 0.00 collisions with false detectors in probe trial 1 and took 3.31 seconds on average. The groups differed significantly in the errors ($z = -2.58$, $p = 0.010$) but not in the IRI variable ($z = -0.67$, $p = 0.50$). In the second probe trial (cue rotation) the two groups did not differ in both of these measures (errors: $z = -1.32$, $p = 0.188$; IRI: $z = -1.90$, $p = 0.058$). Group 1 made an average of 2.56 errors in that trial and took a mean IRI of 7.46 seconds. Participants of group 3 made 8.07 errors in average and achieved an IRI of 12.26 seconds. This surprising finding in the non-parametric tests results from the fact that in group 3 three of the participants scored more than 20 errors in this test leading to the observed numeric difference (Fig. 6a). When the participants were confronted with the third probe trial (cue deletion) group 1 performed with fewer errors ($z = -2.08$, $p = 0.038$) and showed a significantly smaller IRI than group 3 ($z = -3.00$, $p = 0.003$). In that trial group 1 made an average of 0.31 errors and needed 2.52 seconds in the IRI variable, group 3 visited 1.2 false locations and had an IRI of 5.22 seconds on average. After a delay of approximately 30 minutes the last probe trial (delay) started. Within this last probe trial participants of group 1 collided with 3.31 locations on average and took 3.17 seconds in the IRI measure. Group 3 visited an average of 3.36 detectors and needed a mean IRI of 5.26 seconds. Again participants did not differ in the average number of errors ($z = -0.657$, $p = 0.511$) nor the IRI-measure ($z = -1.29$, $p = 0.197$).

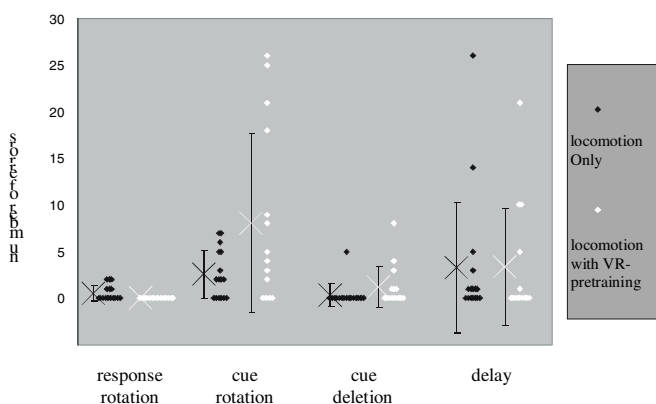
In summary it can be concluded, that it is possible to transfer information that was acquired in a VR environment into an equivalent environment in the real world, although a change of the cue configuration can lead to specific orientation problems in those participants who received the VR pretraining.

4 Discussion

Our goal was to find out whether spatial information that was acquired in a virtual maze can be transferred into the real world. Two groups participated in this experiment. One of the groups was confronted with a virtual maze and after that was transferred into the Kiel locomotor maze. The other group was confronted with the locomotor maze only and could not benefit from a training period. The first question that has to be answered is whether the two versions of the maze show the same degree

of task difficulty. Within the exploration and acquisition phase of the experiment the group in the VR-maze and the other group that was confronted with the locomotor maze did not differ in terms of errors. Therefore we could conclude that the two versions are equal in their degree of difficulty although there seems to be a slight tendency towards a longer exploration and acquisition phase in the VR-maze. This is supported by the observation that participants who were confronted with the VR-maze needed significantly more acquisition trials to achieve the learning criterion, made more errors in the first two trials of acquisition and needed a longer IRI in the exploration and acquisition phase than the group that did the locomotor task. These slight differences in the acquisition phase could be due to the fact that in the VR-world auditory and visual are the only feedback channels the participants can use, whereas in the locomotor task there is a richer supply of multi sensory input that provides the participant with additional information which could help solve the task.

(a)



(b)

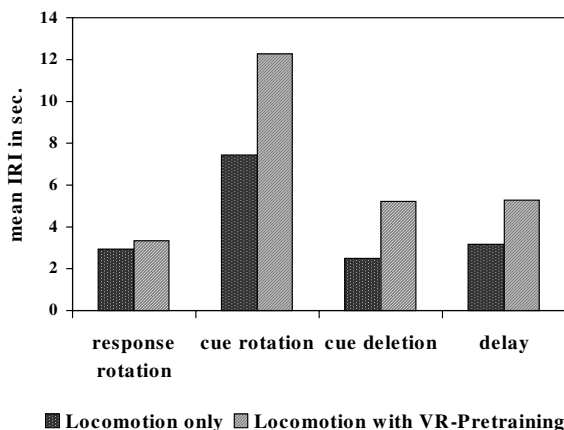


Fig. 6. Plots of individual errors (a) and mean IRIs (b) during probe trials

When the group with the VR-training went into the locomotor maze a clear transfer of the learned knowledge could be observed. Already in the exploration phase of the experiment participants performed less than one error on average, meaning that they were immediately able to see the parallels of the VR-task and the real world task and transfer the learned information. 12 out of 16 participants (75%) of this group performed an exploration trial without any errors. In the acquisition phase these participants also performed significantly better than the other groups. Interestingly enough the IRI measure for this group is significantly higher than the group that was confronted with the locomotor maze only, indicating that the successful transfer of information from a VR to a real world needs a more elaborate or different form of cognitive process than mere exploration behavior reflected by these longer reaction times. Within the acquisition phase the IRI measure of the two groups shows no difference. In this phase it can be assumed that the group with an earlier VR-experience uses the same cognitive processes as the other group. So in spite of the limitations of the desktop VR environment, good quality spatial information can be obtained from this version of the maze.

Within the first probe trial, generally very few errors were made. Still the group that received VR-pretraining made a significantly smaller number of errors than the group that only had to solve the locomotor task. As a matter of fact these participants solved the task without any collision with a false detector indicating that these participants had no trouble with a mental rotation task. It could well be possible that the development of this ability is encouraged more by VR-training than by a training phase in the real world. This view finds its support in the observation made by Larson and colleagues (1999) who found a positive training effect of a virtual reality spatial rotation task in females. After finishing this task females showed an enhanced performance in a paper-and-pencil mental rotation task. The question that arises from these observations is, whether the ability of mental rotation is an important one in order to successfully navigate through a virtual environment. Therefore one could assume that participants who were able to achieve the learning criterion in our task received a generally more intensive training in this aspect of spatial cognition. This is supported by the observation that both groups did not differ in the IRI measure in that probe trial.

For the second probe trial (cue rotation) no significant differences were found either in the IRI measure or the error rates although clear numeric differences can be seen (Fig. 6a+b). This effect results in the fact that very few participants of this group seem to have had problems with the dissociation of cues and therefore scored a considerable number of errors. This result confronts us with the problem that obviously a very small number of people with training of a spatial setup in the VR-environment are largely disturbed in their transfer performance by a dissociation of the presented cues. These findings should be kept in mind when thinking about the possible use of VR-environments as a training means. This idea finds support in the observations made in the third probe trail (cue deletion). Here the differences between the two groups for the error rates and the IRI measure reach a significant level. Again a change of the cue configuration, in this trial the removal of the proximal cues, leads to greater difficulties in solving the task for the participants who received VR-pretraining. Once more the implications for the actual use of the VR-technology as a means of training can not be ignored. If we return to the example mentioned in the introduction, VR-training of orientation on a ship with a tilt would only be a

successful preparation for the case of emergency if possible changes in the environment such as removal of fire extinguishers etc. were integrated in this training.

In the delay phase of the probe trials no differences were observed between the two groups in the error rate and the IRI measure indicating that the participants of both groups were able to retain the previously acquired information over a longer period. In this case the delay-interval was only 30 minutes and the question still remains whether retention of spatial knowledge that is acquired in the VR is comparable to the retention of spatial information that is acquired in the real world over longer periods of time as well. This knowledge could be of greater importance if we consider the training aspects of VR worlds and learning more about this should be kept in mind for further studies.

Concluding the observations of this study we can say that in unimpaired middle-aged adults who were examined, acceptance of this training means was generally good. Although many of the participants were not accustomed to the use of computers and modern computer games like children, who showed that transfer is possible in a previous study (Foreman et al. 2000), from our data it can be concluded that spatial information transfers effectively in healthy adult participants. And yet at this point of research we have to reduce this statement to stable environments. Small changes in the environment seem to have a strong impact on a very small number of participants as could be observed in test 2 and to a larger degree in test 3.

The VR-environment provided in this study does not create such a great sense of immersion as other devices such as head-mounted displays do, but in this environment the problem of cybersickness does not seem to play such an important role as in other studies that use technology with a higher degree of immersion (e.g. Regan & Price, 1994). In addition to this, acceptance of this VR-setup amongst the participants of this study was generally high and therefore this could be a cost effective means for assessing spatial behavior under completely controlled cue conditions. The results could be an indication that the VR-task can be used as a substitute for the locomotor task with the advantage of a transportable task. This new version could be particularly useful for patients with mobility impairments. For the future studies should be aimed at patients with spatial deficits in order to find out how well they cope with the slightly elevated task difficulty of the VR-task that was observed in this study.

The results of this study are a promising approach towards the use of VR-technology in neuropsychological assessment and rehabilitation and give us an impression of how this technology could have an impact on this field in the future.

Acknowledgment

The authors are indebted to Dipl.-Ing. Arne Herzog, an engineer who intensively supported us by working on our hardware and data recording techniques. Also we would like to thank Dipl.-Inf. Lingju Zeng who did the programming and developed the VR-environments. In addition, we wish to thank cand. phil. Ricarda Gross, cand. phil. Mamke Schark, cand. phil. Birgit Heimann, and cand. phil. René Gilster who worked on this project as student research assistants.

References

- Antunano, M. & Brown, J. (1999). The use of Virtual Reality in Spatial Disorientation Training. *Aviation, Space, and Environmental Medicine*, 70(10), 1048.
- Emmett, A. (1994). Virtual reality helps steady the gait of Parkinson's patients. *Computer Graphics World*, 17, 17-18.
- Foreman, N., Stirk, J., Pohl, J., Mandelkow, L., Lehnung, M., Herzog, A., & Leplow, B. (2000). Spatial information transfer from virtual to real versions of the Kiel locomotor maze. *Behavioral Brain Research*, 112, 53-61.
- Johnson, D.A., Rose, F.D., Rushton, S., Pentland, B., & Attree, E.A. (1998). Virtual reality: a new prosthesis for brain injury rehabilitation. *Scottish Medical Journal*, 43(3), 81-83.
- Larson, P., Rizzo, A.A., Buckwalter, J.G., Van Rooyen, A., Krantz, K., Neumann, U., Kesselman, C., Thiebaux, M., & Van der Zaag, C. (1999). Gender issue in the use of virtual environments. *CyberPsychology & Behavior*, 2(2), 113-123.
- Lehrl, S. (1975). Mehrfachwahl-Wortschatztest MWT-B Erlangen: perimed Verlag.
- Leplow, B., Höll, D., Zeng, L., & Mehdorn, M. (1998). Spatial Orientation and Spatial Memory Within a 'Locomotor Maze' for Humans. In Chr. Freksa, Chr Habel and K. F. Wender (Eds.) *Lecture Notes of Artificial Intelligence 1404/ Computer Sciences/ Spatial Cognition*, pp 429-446, Springer: Berlin.
- Leplow, B., Höll, D., Zeng, L., & Mehdorn, M. (2000). Investigation of Age and Sex Effects in Spatial Cognition. In Chr. Freksa, W. Brauer, Chr Habel and K. F. Wender (Eds.) *Lecture Notes of Artificial Intelligence 1849/ Computer Sciences/ Spatial Cognition*, pp 399-418, Springer: Berlin.
- Mehlitz, M., Kleinoeder, T., Weniger, G., & Rienhoff, O. (1998). Design of a virtual reality laboratory for interdisciplinary medical application. *Medinfo*, 9(2), 1051-1055.
- Morris, R.G.M. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation*, 12, 239-260.
- Nelson, H.E. & O'Connell, A. (1978). Dementia: The estimation of pre-morbid intelligence levels using a new adult reading test. *Cortex*, 14, 234-244.
- Reiss, T. & Weghorst, S. (1995). Augmented reality in the treatment of Parkinson's disease. In *Interactive technology and the paradigm for healthcare*. Edited by Morgan, K. Satawa, R.M., Sieburg, H.B., Mattheus, R., Christensen, J.P. Amsterdam IOS Press; 415-422.
- Regan, E., Price, K.R. (1994). The frequency of occurrence and severity of side-effects of immersion virtual reality. *Aviat Space Environmental Medicine*, 65, 527-530.
- Riva, G. (1998). Virtual Environments in Neuroscience. *Transactions on Information Technology in Biomedicine*, 2(4), 275-281.
- Rizzo, A.A. & Buckwalter, J.G. (1997). Virtual reality and cognitive assessment and rehabilitation: the state of the art. *Studies in Health Technology and Informatics*, 44, 123-145.
- Rose, F.D., E.A. Attree, E.A., & Johnson, D.A. (1996). Virtual reality. an assistive technology in neurological rehabilitation. *Current Opinion in Neurology*, 9, 461-467.
- Rose, F.D., Attree, E.A., & Brooks, B.M. (1997). "Virtual environments in neuropsychological assessment and rehabilitation" in *Virtual Reality in Neuro-Psycho-Physiology*, G. Riva (Ed.) Amsterdam, The Netherlands:IOS, 147-156.
- Skelton, R.W., Bukach, C.M., Laurance, H.E., Thomas, K.G., & Jacobs, J.W. (2000). Humans with traumatic brain injuries show place-learning deficits in computer-generated virtual space. *Journal of Clinical and Experimental Neuropsychology*, 22(2), 157-175.
- Thomas, K.G., Hsu, M., Laurance, H.E., Nadel, L., & Jacobs, J.W. (2001). Place learning in virtual space. III: Investigation of spatial navigation training procedures and their application to fMRI and Clinical neuropsychology. *Behavioral Research Methods Instruments, and Computers*, 33(1), 21-37.
- Wilson, B.A., Cockburn, J., & Baddeley, A.D. (1985). *The Rivermead Behavioural Memory Test*. Thames Valley Test Company, Suffolk, England.

Acquisition of Cognitive Aspect Maps

Bernhard Hommel^{1,3} and Lothar Knuf^{2,3}

¹Leiden University, Department of Psychology, Cognitive Psychology Unit,
P.O. Box 9555, 2300 RB Leiden, The Netherlands

Hommel@fsw.LeidenUniv.nl

http://www.fsw.leidenuniv.nl/www/w3_func/Hommel

²Grundig AG Usability Lab, Beuthener Str. 41, 90471 Nuremberg, Germany

Lothar.Knuf@grundig.com

³Max Planck Institute for Psychological Research, Munich, Germany

Abstract. Two experiments investigated the cognitive consequences of acquiring different aspects of a novel visual scene. Subjects were presented with map-like configurations, in which subsets of elements shared perceptual or action-related features. As observed previously, feature sharing facilitated judging the spatial relationship between elements, suggesting the integration of spatial and non-spatial information. Then, the same configuration was presented again but both the features' dimension and the subsets defined by them were changed. In Experiment 1, where all spatial judgments were performed in front of the visible configuration, neither the novel features nor the inter-element relations they implied were acquired. In Experiment 2, where the configurations were to be memorized before the critical judgments were made, novel features were acquired, in part counteracting previous effects of feature overlap. Results suggest that different, subsequently acquired aspects of the same scene are integrated into a common cognitive map.

1 Introduction

Maps are media to represent our environment. They use symbols that are arranged in a particular fashion to represent relevant entities of the area in question and the way these entities are spatially related. However, as maps are not identical with, and not as rich as what they represent they necessarily abstract more from some features of the represented area than from others. For example, a road map contains information that a map of the public transportation network is lacking, and vice versa (Berendt, Barkowsky, Freksa, & Kelter, 1998). Thus, maps are always selective representations of the represented area, emphasizing some aspects and neglecting others.

The same has been shown to be true for cognitive representations of the environment. Far from being perfect copies of the to-be-represented area, cognitive maps often reflect attentional biases, internal correction procedures, and retrieval strategies. As with aspect maps this does not necessarily render them unreliable or even useless, they just do not represent picture-like duplications of the environment but are, in a sense, cognitive aspect maps. Numerous studies provide evidence that cognitive maps are tailored to the needs

and attentional preferences, and sometimes also the cognitive limitations, of their owners (for overview see McNamara, 1991; Tversky, 1981). Our own research has focused onto the role of salient perceptual factors and of action-related information in the processing of visual arrays, such as shown in Figure 1. The most robust finding in several studies was that if people judge the spatial relations between elements of two-dimensional map-like arrays, they are substantially faster if these elements either share a salient perceptual feature, such as color or shape (Gehrke & Hommel, 1998; Hommel, Gehrke, & Knuf, 2000), or if they have been learned to signal the same action (Hommel & Knuf, 2000; Hommel, Knuf, & Gehrke, 2002). Moreover, these effects are independent of whether the judgments are given in front of a novel array or made from memory, ruling out factors having to do with memory organization, retrieval, or selective forgetting. Rather, perceptual or action-related commonalities between elements seems to induce the formation of cognitive clusters connecting the representations of the related elements via the shared feature code (Hommel & Knuf, 2000). Accordingly, accessing the codes of one element spreads activation to connected elements, thereby facilitating comparison processes. That is, people acquire cognitive maps the structure of which represented one particular, salient aspect of the to-be-represented environment—hence, cognitive aspect maps.

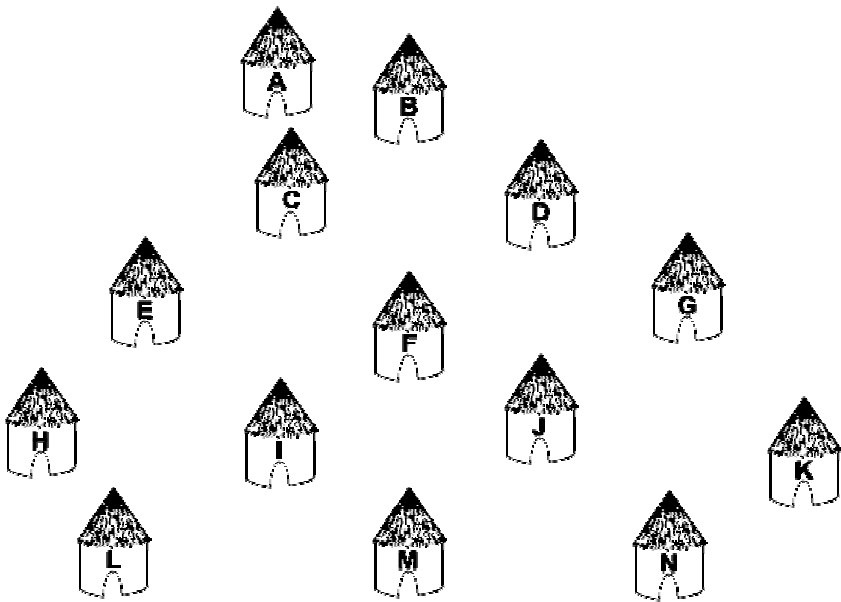


Fig. 1. Example of the stimulus layout used in all experiments. The huts were displayed at nearly the same locations for each participant, only differing by a small jitter of up to 5 cm per location (to counteract possible emerging figural properties of the display). The letters indicating the locations were not shown to the subjects; instead each hut was identified by a nonsense “name” (i.e., a meaningless syllable like “MAW”, omitted here) appearing at the letter’s position. Note that the hut in a particular location had a different name for each participant.

Previous studies were restricted in that they introduced only one dimension of similarity or feature sharing at a time, that is, there was only one salient aspect of the array. Yet, in everyday life we are often confronted with alternative aspects of the same environment. For instance, we go walking, ride a bike, take a subway, or drive by car in the same city, thereby following different tracks and routes, observing different constraints and, hence, focusing on different aspects of the same area. How are these different aspects cognitively represented? One possibility, suggested by computational approaches to aspect-map representation (e.g., Berendt et al., 1998) were to acquire and store independent cognitive maps and to retrieve them according to the current task and goal. Alternatively, people may begin with forming a cognitive map with respect to one aspect and fill in additional information, such as new links between locations, when focusing on another aspect (e.g., McNamara & LeSueur, 1989). That is, the same cognitive map may be used to represent all the acquired aspects—which may be differentially marked to relate them to the relevant aspect.

Importantly for both psychological and empirical reasons, the separate-maps and the integrative-maps view differ in their predictions with respect to the effect of acquiring information about a new aspect of an already known array. According to the separate-maps view there is no reason to assume that learning about aspect B of a given array X would change the representation of X with respect to another aspect A. Both aspects should be stored in different cognitive maps which should not interact. According to the integrative-maps view, however, learning about B should indeed be suspected to modify the map, especially if the implications of aspect B contradict the implications of aspect A. For example, assume subjects acquire a visual array as depicted in Figure 1. Assume that in a first trial the huts labeled B and F are presented in the same color, whereas F and M appear in different colors. If subjects would then verify spatial relations between hut pairs they should perform better when comparing B and F than when comparing F and M, indicating that perceptual grouping by color induced the creation of corresponding cognitive clusters. However, what would happen if, in a second trial, F and M were mapped onto the same response, while B and F required different responses (a condition that we know to induce action-based cognitive clustering)? This would change the similarity relationship between the three items: B and F would be alike with respect to one aspect but different with respect to another, and the same were true for F and M. Hence, the huts would be parts of aspect relations that are, in a sense, incongruent with each other.

According to the separate-maps approach, introducing different (and presumably differently-clustered) aspects would be expected to lead to the acquisition of two different cognitive aspect maps. If so, one map were used to perform in one part of the task and another map in the other part, so that the effects of inter-item similarity should be independent; i.e., subjects should perform better on B-F in the color condition and better on F-M in the action condition. According to the integrative-maps view, however, different aspects are integrated into the same cognitive map, so that learning about a new aspect might affect performance on the items in question. In our example, having learned that B and F are alike with respect to one aspect might facilitate comparing B and F even if, in the following, subjects learn that B and F are dissimilar regarding another, new aspect. If so, color-based similarity and action-based similarity would work against each other, which should decrease the effect of action-based similarity as compared to a condition where this type of similarity is acquired first. Inversely, later tests of the effect of color-

based similarity should be reduced by exposure to the differing action-based similarity. Whether this is so we tested in two pairs of experiments.

2 Experiment 1

In Experiments 1A and 1B, subjects judged spatial relationships between houses of an imaginary village arranged as in Figure 1. All judgments were carried out vis-à-vis the visual array, hence, the task was purely perceptual in nature. Each of the two experiments 1A and 1B consisted of three blocks. The first blocks were closely modeled after our previous studies, where we found comparison speed to be affected by inter-item similarity based on color (Gehrke & Hommel, 1998; Hommel et al., 2000) and shared action (Hommel & Knuf, 2000; Hommel et al., 2002)—which we take to imply color- and action-induced cognitive clustering. That is, in Experiment 1A the houses of our imaginary village looked all the same except that they were colored in such a way that three (configuration C3) or four (C4) color groups were formed. Correspondingly, in Experiment 1B subjects learned that the houses were mapped onto three (C3) or four (C4) keypressing actions. On the basis of our previous findings we expected the time needed to verify a statement regarding the spatial relation of two given houses to be less if the two items share the color (in 1A) or action (1B) as if they do not.

In a second block we introduced a new aspect. In Experiment 1A the houses were no longer colored but now required particular keypressing actions. The configuration was changed from C3 to C4, or vice versa, so that the similarity relations implied by color and action agreed in some cases but not in others (B, F, and M). The crucial question was whether similarity effects would be as in the first block of Experiment 1B (where action served to induce similarity as well) or whether they would be affected by previously learning another aspect. Of special diagnosticity for this question was performance on B-F and F-M, the pairs with differing (*incongruent*) similarity relations in the blocks of the experiment. Analogously to Experiment 1A, 1B no longer required particular actions related to houses but introduced new color relationships as in the first block of 1A. Accordingly, the question was whether this would lead to performance equivalent to the first block of 1A, or whether some impact of previously learning another aspect in the first block would show up.

In the concluding third block of the experiments the first condition was rerun (ABA design). Here we were interested to see whether performance would be comparable to the first block, which would suggest that the two acquired aspects are stored in separate, non-interacting maps, or whether after-effects of learning about another aspect in the second block could be demonstrated, as the integrative-maps view suggests.

Apart from the relation-judgment task we also asked subjects to estimate Euclidean distances between pairs of objects. Although distance estimations and the verification of spatial relations are commonly thought to tap into the same cognitive processes, our previous studies consistently revealed a dissociation between these two measures. In particular, we did not obtain any hint that inter-item similarity affects distance estimation. In our view, this suggests that similarities affect the way information of spatial layouts is cognitively organized (a factor that impacts verification times) but not the quality of spatial representations itself, an issue we briefly get back to in the General Discussion. Accord-

ingly, we did not expect interesting effects to show up in distance estimations (and, indeed, there were no such effects) but did include this task in Experiment 1 anyway just to be sure.

2.1 Method

Thirty-five naive male and female adults (mean age 24.5 years) were paid to participate; 23 took part in Experiment 1A, 12 in Experiment 1B. Stimuli were presented via a PC-controlled video beamer on a 144 x 110 cm projection surface, in front of which subjects were seated with a viewing distance of about 200 cm. They responded by pressing different arrangements of sensor keys with the index finger (see below).

Stimuli were map-like configurations of 14 identically shaped houses, appearing as a virtual village (see Figure 1). Houses were displayed at nearly the same locations for each participant, only differing by a small jitter of 5 cm at maximum on each location (to avoid systematic spatial Gestalt effects). They were 15 x 15 cm in size and labeled by consonant-vocal-consonant nonsense syllables without any obvious phonological, semantic, or functional relations to each other or to location-related words—to exclude any cognitive chunking based on house names. The name-to-house mapping varied randomly between subjects.

Table 1. Design of Experiments 1 and 2. Experimental blocks differed in terms of grouping modality (i.e., houses were similar or dissimilar in terms of color or assigned action) and configuration (C3: three different colors or actions; C4: four different colors or actions; see Figure 2). Both modality and configuration alternated from block to block (C3→C4→C3 or C4→C3→C4).

Block	Experiments 1A and 2A		Experiments 1B and 2B	
	Modality	Configuration	Modality	Configuration
1	color	C3 / C4	action	C3 / C4
2	action	C4 / C3	color	C4 / C3
3	color	C3 / C4	action	C3 / C4

The experiment consisted of one experimental session of about 90 min., which was divided into three blocks differing in modality of grouping (Experiment 1A: color → action → color; 1B: action → color → action) and configuration sequence (C3/C4 vs. C4/C3), see Table 1. In the first block of Experiment 1A groupings were induced by color. In configuration C3, three different colors were used to induce three perceptual groups (group C3₁: B, C, D, F; group C3₂: E, H, I, L; and group C3₃: G, J, K, M, N; see Figure 2). In configuration C4, four colors were used to induce four groups (group C4₁: B, C, D; group C4₂: E, H, L; group C4₃: G, K, N; and group C4₄: F, I, J, M). The house in location A always served as neutral item; its only use was to avoid possible end or anchor effects on relation-judgment or estimation performance.

In the second block of Experiment 1A, color was removed from the objects, i.e., the homogenous stimulus layout shown in Figure 1 was presented. Also, the configuration was changed; i.e., subjects confronted with C3 in the first block were now confronted with C4 and vice versa (see Figure 2). Yet, the *spatial* stimulus arrangement for a given

participant remained unchanged throughout the whole experiment. In contrast to the first block, subjects were to perform simple keypressing responses to induce cognitive clusters. In each trial, one of the houses would flash in red and the subject would press one of three or four response keys. The key-to-house mapping varied randomly between participants. As it was not communicated, they had to find out the correct mapping by trial and error. In case of a correct response the (red) color of the current object vanished and the next one was flashed. In case of an error an auditory feedback signal appeared and a different key could be tried out. Once subjects produced correct consecutive responses to all locations in a sequence, the mapping-induction phase ended. The third block was always exactly the same as the first one, i.e., groupings were induced by color and with the same, original configuration.

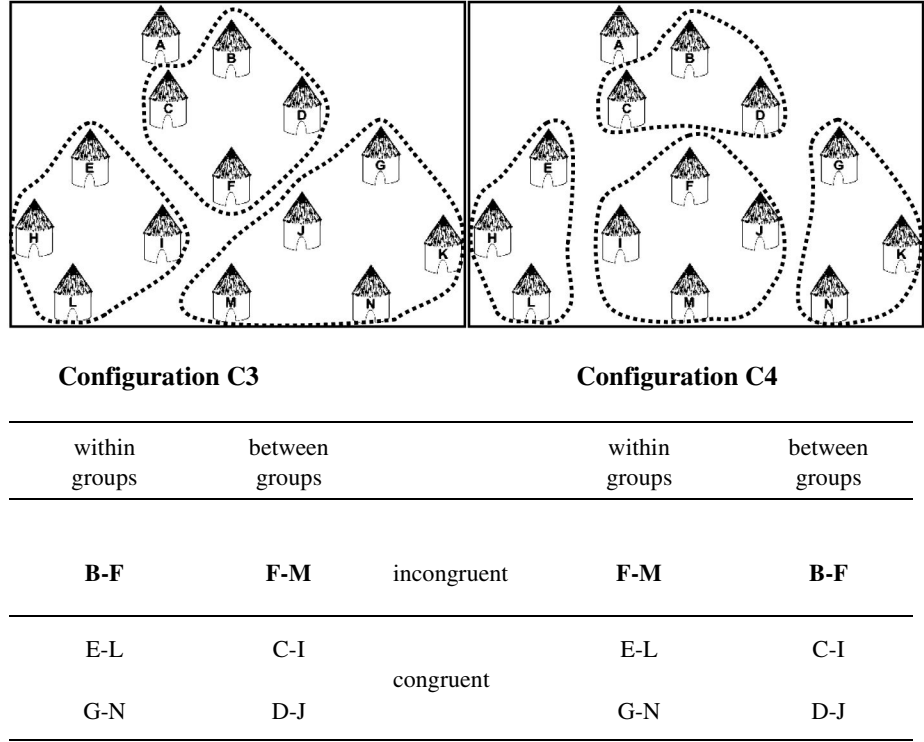


Fig 2. Illustration of groupings by color and actions. Three to five of the huts were either displayed in the same color or assigned to the same keypressing response (groupings or assignments indicated by line borders, which were not shown in the experiments), this making up either three or four perceptual/action-related groups (C3 and C4). The sequence of configurations was always alternated between blocks (C3/C4/C3 vs. C4/C3/C4), as indicated in Table 1. As a consequence, the group membership of the location pairs B-F and F-M changed from block to block. The tables at the bottom indicate which comparisons entered the analyses of group-membership and congruency effects.

In Experiment 1B the method was exactly the same, only that the sequence of color and action blocks was interchanged (action $\rightarrow$ color $\rightarrow$ action).

In each experimental block subjects performed a relation-judgment task and a distance-estimation task in front of the visible stimulus configuration, task order being balanced across subjects. Six vertical location pairs were chosen for distance estimations and relation judgments, each pair being separated by ca. 300 mm. Half of the pairs were composed of houses within the same color or action group and the other half consisted of houses from different groups. In configuration C3, the pairs B-F, E-L, and G-N were assigned to the same color/key, while the pairs C-I, D-J, and F-M were assigned to different colors/keys (see Figure 2). In configuration C4, the respective within-group pairs were F-M, E-L, and G-N and the between-group pairs C-I, D-J, and B-F. As configurations varied between blocks (i.e., C3 $\rightarrow$ C4 $\rightarrow$ C3 or C4 $\rightarrow$ C3 $\rightarrow$ C4, see Table 1), group membership of some location pairs changed from 'between' to 'within' and vice versa. Those critical, *incongruent* location pairs were B-F and F-M.

Distance Estimations. Thirty-six critical pairs of house names (3 repetitions of the 6 critical pairs presented in the 2 possible orders) and 12 filler pairs were displayed, one pair at a time, in the upper center of the projection surface. The names were displayed in adjacent positions, separated by a short horizontal line, serving as hyphen. Another horizontal line of 70 cm in length was shown above the names and participants were explained that this line would represent 150 cm (more than the width of the whole projection surface). It was crossed by a vertical pointer of 5 cm in length, which could be moved to the left or right by pressing the left and right response key, respectively. For each indicated pair, participants were required to estimate the distance between the corresponding objects (center to center) by adjusting the location of the pointer accordingly, and then to verify their estimation by pressing the two response keys at the same time.

Relation Judgments. On basis of the 6 critical items a set of 128 judgments was composed, consisting of 4 repetitions for each item, 2 relations (*under*, *above*), and 2 presentation orders (*A-relation-B*, *B-relation-A*). 32 judgments on distractor pairs were added to the set. The to-be-verified relation statements were presented one at a time. In each trial, a fixation cross appeared for 300 ms centered on the top of the display. Then the statement appeared, consisting of the names of two objects and a relation between them, such as "RUK under JOX" or "KAD above NOZ". Participants were instructed to verify the sentence as quickly and as accurately as possible by pressing the 'yes' or 'no' key accordingly, assignment of answer type and response key being counterbalanced across participants. The sentence stayed on the projection surface until response. After an inter trial interval of 1000 ms the next trial appeared. In case of an incorrect keypress an error tone appeared and the trial was repeated in a random position within the remaining series of trials. If the same error on the same trial was made for three times, this trial was excluded from the data.

2.2 Results and Discussion

Data were coded as a function of experimental block (1-3), group membership (*within-group* vs. *between-group*) and congruency (*congruent* vs. *incongruent*), as indicated in the scheme shown in Figure 2 (bottom). Thus, performance on distractor pairs was not analyzed. Analyses employed a four-way mixed ANOVA with the within-subjects factors group membership, congruency, and experimental block, and the between-subjects factor experiment (1A vs. 1B). The significance level was set to $p < .05$ for all analyses.

From the data of the *distance-estimation task*, mean estimates in millimeters were computed. Across all conditions, the real distance of 300 mm was underestimated (Mean = 215 mm, SD = 46 mm). However, the ANOVA did not reveal any reliable effect or interaction, suggesting that there were no systematic distortions for object pairs spanning one vs. two groups, or for congruent vs. incongruent relations.

In the *locational-judgment task*, error rates were below 2% and the respective trials were excluded from analysis. The four-way ANOVA revealed a highly significant main effects of experiment, $F(1,22) = 16.862$, showing that RTs were generally slower in Experiment 1A than in 1B, and of block, $F(2,44) = 242.312$, indicating a decrease of RTs across blocks. More importantly, a highly significant main effect of group membership was revealed, $F(1,22) = 22.027$, indicating that relations between objects of the same color or action group were verified faster than relations between objects of different groups. However, this effect was modified by an interaction of group membership and block, $F(2,44) = 4.860$, indicating that grouping effects were reliable in Block 1 and 3, but not in Block 2. This effect was not further modulated by experiment ($p > .9$), suggesting that the way how groupings were induced did not play a role.

A main effect of congruency was also obtained, $F(1,22) = 18.922$, showing *slower* RTs for congruent object pairs than for incongruent ones. On first sight, this is a counter-intuitive effect—it not only goes in the wrong direction, it also suggests that subjects were able to anticipate in the first block already which locations were rendered congruent or incongruent by the changes in the second block. Yet, note that it was always the same spatial locations that were used for the congruency manipulations (locations B, F, and M). Accordingly, a main effect of congruency merely reflects the relative difficulty to process information from these locations. As they occupied the horizontal center of the display, they may have been more difficult to find than more peripheral locations and/or processing the items presented there have suffered from the relatively high degree of masking from surrounding items.

At any rate, the more interesting question was whether grouping effects behaved differently for congruent and incongruent items. Indeed, besides an interaction with block, $F(2,44) = 7.547$, and with block and experiment, $F(2,44) = 3.531$, congruency entered a triple interaction with group membership and block, $F(2,44) = 4.925$; all further interactions failed to reach significance. To decompose the latter effect, separate ANOVAs were computed for congruent and incongruent trials. As suggested by Figure 3, no interaction effect was obtained for congruent trials. However, for incongruent trials group membership interacted with block, $F(2,44) = 6.989$, due to that standard grouping effects occurred in the first and the third block, but were reversed in the second block. As the status of within- and between-groups pairs changed under incongruence, this means that the original grouping effect from the first block persisted in the second block. In other words,

subjects did not react to the grouping manipulation in the second block. (Indeed, membership no longer interacted with block when we reversed the sign of group membership in Block 2, that is, when we determined group membership for items in all blocks on the basis of their membership in Block 1.) As the critical interaction was not modified by experiment ($p > .9$), this lack of an effect can not be attributed to the way grouping was induced. Indeed, a look at the results from the first blocks shows that substantial grouping effects were induced by both color and action manipulations. Hence, commonalities with respect to both color and action seem to induce comparable cognitive clusters, but only if they are present the first time the stimulus configuration is encountered. Once the clusters are formed, so it seems, shared features are ineffective. In other words, acquiring one cognitive aspect map of an array blocks the acquisition of another aspect map.

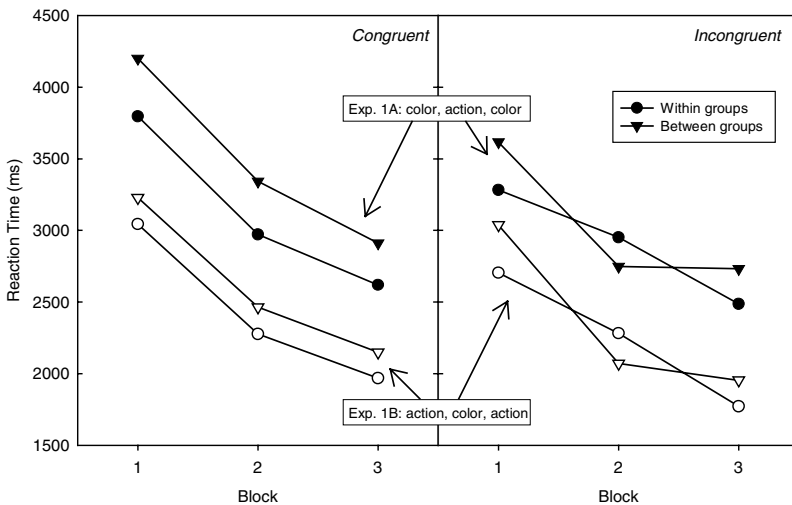


Fig. 3. Mean reaction times for verifying spatial relations between pairs of elements belonging to the same (within groups) or different (between groups) color- or action-induced group, as a function of block. Black symbols refer to Experiment 1A, white symbols to Experiment 1B.

To summarize, we see that the acquisition of perceptual or action-related aspects of a visual array strongly depend on previous experience. In particular, having experienced that the array items are similar with respect to one aspect—be it perceptual or action-related—prevents any further effect of other types of similarity. On the one hand, this is indicated by the fact that facing a novel aspect that supports an already acquired similarity relation, such as when items shared both color and action, does not increase the grouping effect. That is, in the left, congruency panel of Figure 3 there is not the slightest hint to an increase of the grouping effect in blocks 2 and 3 as compared to block 1, and this is so for 1A and 1B. On the other hand, there is also no hint to any grouping effect of the novel aspect in the incongruency condition. On the contrary, the pattern shown in the right, incongruency panel of Figure 3 shows that the grouping effect in block 2 entirely follows the old grouping encountered in the first block but shows no sign of effect by the present

grouping. And finally, performance in the third block pretty much mirrored that in the first block, suggesting that the intermediate experience with another aspect had no effect.

3 Experiment 2

The outcome of Experiment 1 suggests that having structured a novel visual array with regard to one perceptual or functional dimension kind of immunizes the perceiver/actor against alternative ways to structure that array. It is as if perceivers/actors search for some obvious characteristic of the to-be-processed scene suited to provide the basic, internal structure of the scene's cognitive representation, and once a satisfying characteristic has been identified no other is needed. Yet, the situations in which we introduced and offered new features to induce some re-structuring of our subjects' scene representations were not too different from the previous ones and the tasks the subjects solved were rather similar. Hence, there was no real reason or motivation for subjects to re-structure their cognitive maps, so that our test for re-structuring effects was arguably weak. Moreover, all data we obtained were from purely perceptual tasks that, in principle, could be performed without any contribution from higher-level cognitive processes. Hence, our tasks arguably minimized, rather than maximized chances to find contributions from such cognitive processes.

Experiment 2 was carried out to provide a stronger test. Rather than merely confronting subjects with the visual arrays and asking them to carry out relation judgment we from Block 2 on required them to make these judgments from memory. In particular, we in Block 1 induced groupings by color (in Experiment 2A) or shared action (in Experiment 2B) and asked subjects to perform relation judgments in front of the visual array, just like in Experiment 1. Then, in Block 2, we introduced shared action or color as second grouping dimension, respectively, but here subjects were first to learn the spatial array before then making their judgments from memory. In Block 3 we switched back to the grouping dimensions used in Block 1 and tested again from memory. These design changes were thought to motivate subjects to establish new cognitive maps, or at least update their old ones, in Block 2 and, perhaps, in Block 3 as well. If so, we would expect an increasing impact of incongruent groupings in Block 2 and, perhaps, some impact on performance in Block 3.

3.1 Method

Twenty-four adults (mean age 23.1), 12 in Experiment 2A and 12 in 2B, were paid to participate. Apparatus and stimuli were the same as in Experiment 1, as was the sequence of blocks.

In contrast to Experiment 1, however, the mapping induction by keypressing responses in the second block of Experiment 2A was followed by an active learning phase. Following a 2-min study period, the configuration disappeared and the participants were sequentially tested for each object. A rectangle of an object's size appeared in the lower right corner of the display, together with an object name in the lower left corner. Using the same keyboard as before, participants moved the rectangle to the estimated position of

the named object and confirmed their choice by pressing the central key. Then the projection surface was cleared and the next test trial began. There were 14 such trials, one for each object, presented in random order. If in a sequence an object was mislocated for more than about 2.5 cm, the whole procedure was repeated from the start. The learning phase ended after the participant completed a correct positioning sequence.

Thereafter the mapping induction was repeated to prevent decay of information about the house-key mapping (Hommel et al., 2002). Since the stimulus layout was no longer visible, the name of a house appeared on the top of the screen and the correct key-to-house mapping had either to be recalled or again found out by trial and error. After having acquired the valid house-key mappings, subjects verified sentences about spatial relations between houses from memory. Distance estimations were not obtained.

Block 3 was also performed under memory conditions, so color-based grouping had to be reintroduced. The configuration of colored objects was therefore shown for about 2 minutes at the beginning of a new acquisition phase as well as at the beginning of each positioning sequence (see below). The rest of the procedure followed Experiment 1. Experiment 2B differed from 2A only in the sequence of grouping types (action → color → action) and was therefore a replication of Experiment 1B under mixed perceptual and memory conditions.

3.2 Results and Discussion

A four-way mixed ANOVA of verification times revealed a significant main effect of experimental block, $F(2,44) = 68.562$, indicating that RTs decreased across blocks (see Figure 4). This practice effect was more pronounced in Experiment 2A, which produced a block $\times$ experiment interaction, $F(2,44) = 3.807$. A main effect of congruency was obtained, $F(1,22) = 5.487$; it was again negative, showing slower RTs for congruent than

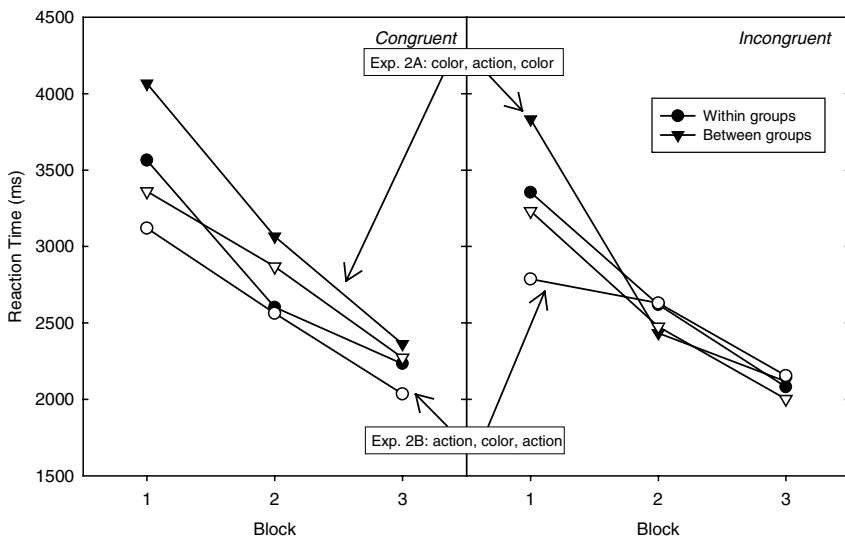


Fig. 4. Mean reaction times for verifying spatial relations between pairs of elements belonging to the same (within groups) or different (between groups) color- or action-induced group, as a function of block. Black symbols refer to Experiment 2A, white symbols to Experiment 2B.

incongruent pairs, and therefore is likely to reflect the general difficulty to process information from central locations.

More importantly, a highly significant main effect of group membership was obtained, $F(1,22) = 18.493$, indicating that relations between objects of the same color or action group were verified faster than relations between objects of different groups. This effect was modified by a group membership $\times$ block interaction, $F(2,44) = 4.408$, and a triple interaction of congruency, group membership, and block, $F(2,44) = 3.449$. Interestingly, these interactions did not depend on the experiment ($p > .9$). As shown in Figure 4, different grouping effects were obtained in Blocks 2-3 than in the first blocks of congruent and incongruent conditions.

In the first blocks of both experiments and under both congruency conditions grouping effects very much like in Experiment 1 were obtained. That is, both shared color and shared action facilitated the judgment of the spatial relations between object pairs to a comparable and replicable degree. In Block 2 the picture changes dramatically. Under congruency, the results again look very much like in Experiment 1, that is, grouping effects are pronounced in all three blocks and (statistically) unaffected by the block factor. Incongruency yielded a different pattern. The second block led to a reversal of the membership effect similar to Experiment 1, but now it was clearly reduced in size and no longer reliable (as revealed by *t*-tests, $p > .05$). The third block behaved quite differently than in Experiment 1. Rather than showing the same sign and size as in Block 1, here the membership effect more or less disappeared ($p > .05$). Thus, the two reversals of group membership in the second and third block clearly affected performance, suggesting that our memory manipulation was effective, indeed.

4 General Discussion

The guiding question of the present study was whether encountering information about a new aspect of an already known visual array leads the creation of a new cognitive aspect map that is stored separately from the original one, or whether the new information is integrated into the original cognitive map, thereby updating and transforming it. According to the separate-map view map acquisition should be unaffected by previously acquired knowledge and cognitive maps created thereof. From this view we would have expected that congruency between acquired and novel aspects has no impact on map acquisition, so that in Experiment 1 performance in congruent and incongruent conditions of Block 2 should have been comparable. However, performance clearly differed, in that novel aspects were not acquired if the grouping they implied was incongruent with the grouping induced by previous experience. In fact, previous experience with one group-inducing aspect seemed to have completely blocked out any effect of a novel aspect, so that performance in Block 2 perfectly matched performance in Block 1.

These results rule out the separate-maps approach, as it is unable to account for interactions between cognitive maps or side-effects of already existing maps. However, the findings are also inconsistent with the integrative-maps approach in demonstrating that new information was simply not integrated. Apparently, when encountering a new visual array people spontaneously pick up actually irrelevant features shared by subsets of its elements to create a clustered cognitive map; yet, once a map is created it does not

elements to create a clustered cognitive map; yet, once a map is created it does not seem to be spontaneously updated. However, the findings obtained in Experiment 2 suggest that updating does take place when people are given a reason to modify their cognitive maps. Not only is new information acquired under these conditions, it is also integrated into the existing cognitive map, as indicated by the disappearance of the membership effect under incongruency in Block 2 and 3. Thus, we can conclude that people do not under all circumstances store the aspects of a visual scene they come across, but if they do so they integrate them into a single, coherent cognitive map. This insight, together with the result pattern of the present study, has several implications, three of which we will discuss in turn.

4.1 Representing Aspect Maps

A first, theoretical implication relates to how spatial arrays are cognitively represented. Commonly, effects of nonspatial properties on spatial representations are taken to imply some kind of hierarchical representation, in which spatial information is stored within nested levels of detail with levels being organized by nonspatial categories (e.g., McNamara, 1986; McNamara, Hardy, & Hirtle, 1989; Palmer, 1977). To support such hierarchical representations authors often refer to known memory distortions, such as the relative underestimation of distances between cities belonging to the same state (e.g., Stevens & Coupe, 1978).

However, as we have pointed out elsewhere (Hommel & Knuf, 2000) effects of nonspatial relations on spatial judgments can be understood without reference to hierarchies. Consider the cognitive architecture implied by our present findings. Figure 5 shows an account of these findings along the lines of TEC, the *Theory of Event Coding* proposed by Hommel, Müsseler, Aschersleben, and Prinz (in press; Hommel, Aschersleben, & Prinz, in press). TEC makes two assumptions that are crucial for our present purposes. First, it assumes that perceived events (stimuli) and produced events (actions) are cognitively represented in terms of their features, be they modality-specific, such as color, or modality-independent, such as relative or absolute location. Second, TEC claims that perceiving or planning to produce an event involves the integration of the features coding it, that is, a binding of the corresponding feature codes.

Figure 5 sketches how these assumptions apply to our present study. Given the features each hut possessed in our study, its cognitive representation is likely to contain codes of its name, location, color, and the action it requires (cf., Hommel & Knuf, 2000). As TEC does not allow for the multiplication of codes (i.e., there is only one code for each given distal fact), sharing a feature implies a direct association of the corresponding event representations via that feature's code. That is, if two huts share a color or an action, their representations include the same feature code, and are therefore connected. Along these connections activation spreads from one representation to another, so that judging the relation between objects that have associated representations is facilitated. In congruent cases (i.e., if the current association is compatible with previously acquired associations) activation spreads to representations of only those objects that currently share some aspect (see panel A). However, in incongruent cases activation spreads to both objects currently sharing an aspect and objects that previously shared some aspect (see panel B).

As a consequence congruent, but not incongruent cases give rise to standard group-membership effects, just as observed in the present study.

Interestingly, along these (non-hierarchical) lines category-induced effects on spatial judgments can be explained as well. Consider, for instance, the three huts depicted in Figure 5 were all of the same color and not associated with different actions, but DUS and FAY were known to belong to a hypothetical "County A" while MOB belonged to "County B" (the category manipulation used by Stevens & Coupe, 1978). According to TEC, such a category membership is just another feature that, if it code is sufficiently activated and integrated, becomes part of the cognitive representation of the respective hut. Thus, instead of the code "red" or "green" the representations of DUS and FAY would contain the feature code "County A member", whereas the representation of MOB contained the code "County B member". If so, DUS and FAY were associated the same way as if they were of the same color, so that judging their spatial relation would be faster than judging that between FAY and DUS. Hence, category effects do not necessarily imply hierarchical representations but may be produced the same way as effects of perceptual or action-related similarities.

From comparing the outcomes of Experiments 1 and 2 it is clear that the when and how of feature integration depends on the task context. The results of Experiment 1 suggest that after having integrated the features available in the first block, subjects did not continuously update their event representations but went on operating with the already acquired ones. Accordingly, the new features introduced in the second block were not considered, their codes were not integrated, and therefore did not connect the representations of the objects sharing the particular feature. In contrast, asking subjects to memorize the display in Experiment 2 seems to have motivated (or even required) the update of the object representations, which provided a chance for the new features to get integrated. Thus, although the selection of features to be integrated does not seem to be determined intentionally (as indicated by color- and action-induced effects), the timepoint or occasion of integration is.

4.2 Assessing Aspect Maps

A second implication of our findings refers to method. Many authors have taken the speed of spatial judgments and distance estimations to reflect the same cognitive processes or structures and, hence, to measure the same thing. Yet, in our studies, including the present one, we consistently observed a dissociation between these measures, that is, systematic effects of grouping manipulations on reaction times of relation judgments but not on distance estimations (Gehrke & Hommel, 1998; Hommel et al., 2000, 2002; Hommel & Knuf, 2000). Although accounts in terms of strategies and differential sensitivity are notoriously difficult to rule out, we think it is worthwhile to consider that these measures reflect different cognitive functions. Along the lines of McNamara and LeSueur (1989) it may be that nonspatial information supports (or hinders) particular ways to cognitively structure information about visual scenes (assessed by the speed of comparative judgments) but does not modify its spatial content (assessed by distance estimations). In other words, feature sharing may affect the (ease of) access to cognitive codes but not what these codes represent.

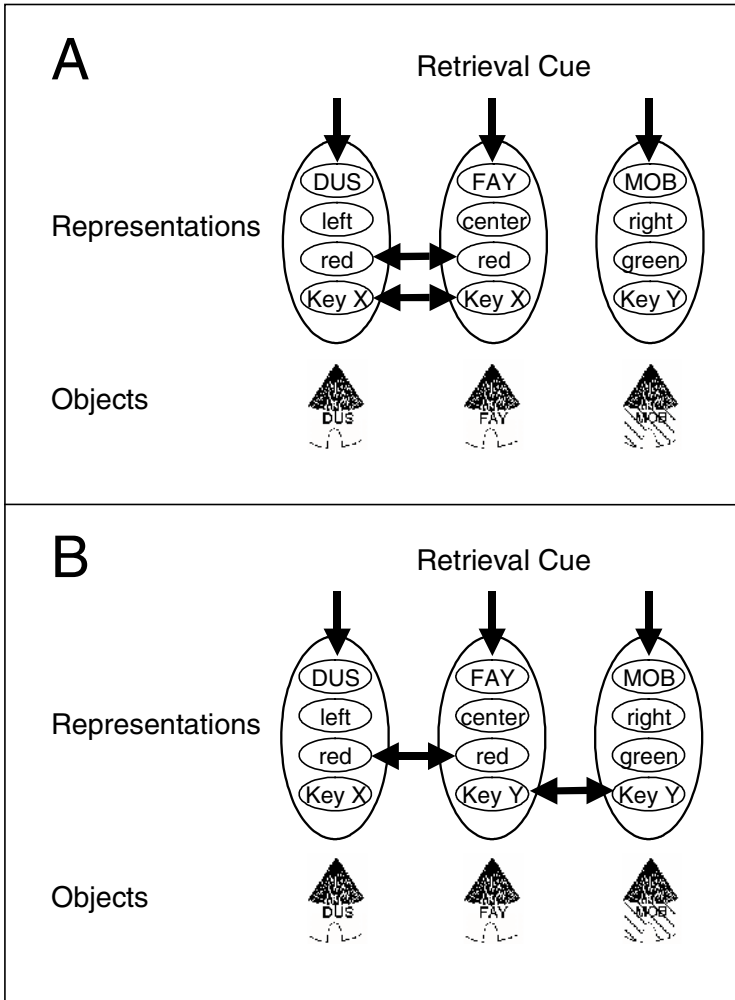


Fig. 5. A simplified model of how feature overlap between elements of a scene may affect the speed of verification judgments. Panel A shows an example of *congruent* learning, in which the hut FAY shared its color with DUS but not MOB on one occasion, and shared an action (response key) with DUS but not MOB on another occasion. This results in a strong association between the representations of DUS and FAY, so that activating the representation of FAY (e.g., in the course of retrieval) spreads activation to DUS, and vice versa. Panel B shows an example of *incongruent* learning, in which FAY shared its color with DUS but not MOB on one occasion, and shared an action with MOB but not DUS on another occasion. As a consequence, FAY becomes associated with both DUS and MOB, so that activating the representation of FAY spreads activation to both DUS and MOB.

Of course, this raises the question why other authors did find distortions of the content of spatial memories (e.g., Stevens & Coupe, 1978; Thorndyke, 1981; Tversky, 1981). We could imagine two types of causes that may underlie such findings. One is configura-

tional, that is, purely visual factors—such as Gestalt laws—may distort the processed information during pick up, so that the memories would be accurate representations of inaccurately perceived information (Knuf, Klippel, Hommel, & Freksa, 2002; Tversky & Schiano, 1989). Another factor relates to response strategies. In many cases it may simply be too much to ask for precise distance estimations because the needed information is not stored. Under decision uncertainty people are known to employ "fast and frugal heuristics" (Gigerenzer & Todd, 1999), so that subjects may use the presence or absence of nonspatial relations, or the degree of mutual priming provided thereby, to "fine-tune" their estimations. How strongly this fine-tuning affects and distorts distance estimations is likely to vary with the degree of uncertainty, which may explain why distortions show up in some but not in other studies.

4.3 Acquiring Aspect Maps

A third implication of our findings is of a more practical nature. There is a growing number of demonstrations in the literature that humans fall prey to all sorts of biases and distortions when forming cognitive maps of their environment—even though we ourselves were unable to find such qualitative effects. Considering these observations one is easily led to adopt a rather pessimistic view on the quality and reliability of spatial representation in humans. However, the present findings suggest that biases and distortions are prevalent only in the beginning of forming a cognitive representation of a novel scene or array. Thus, if we create a new cognitive map we are attracted to and guided by only a few, currently relevant aspects of the represented environment, which is likely to induce one or another distortion under conditions of high decision uncertainty. However, with changing interests, tasks, and ways to get in touch with that environment information about additional aspects will be acquired and integrated into the same cognitive map. By integrating different aspects their possibly biasing and distorting effects will cancel out each other, the more likely the more aspects get integrated. Accordingly, rather than multiplying biases and distortions enriching one's cognitive map will lead to a more balanced, and therefore more reliable spatial representation.

4.4 Conclusion

To conclude, our findings suggest that when people create a cognitive map they are spontaneously attracted by perceptual features and actions (i.e., aspects) shared by subsets of the represented environment, and the way they organize their cognitive maps reflects these commonalities. However, once a scene is cognitively mapped novel aspects are acquired only if there is some necessity, such as posed by requirements of a new task. In that case the new information is integrated into the already existing cognitive representation, thereby modifying its behavioral effects. Hence, features of and facts about our spatial environment are not stored in separate aspect maps but merged into one common map of aspects.

Acknowledgments

The research reported in this paper was funded by a grant of the German Science Foundation (DFG, HO 1430/6-1/2) and supported by the Max Planck Institute for Psychological Research in Munich. We are grateful to Edith Mueller, Melanie Wilke and Susanne von Frowein for collecting the data.

References

- Berendt, B., Barkowsky, T., Freksa, C., & Kelter, S. (1998). In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge* (pp. 313-336). Berlin: Springer.
- Gehrke, J., & Hommel, B. (1998). The impact of exogenous factors on spatial coding in perception and memory. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge* (pp. 64-77). Berlin: Springer.
- Gigerenzer, G., & Todd, P. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. Todd and the ABC research group (Eds.), *Simple heuristics that make us smart* (pp. 3-36). Oxford: University Press.
- Hommel, B., Aschersleben, G., & Prinz, W. (in press). Codes and their vicissitudes. *Behavioral and Brain Sciences*, 24.
- Hommel, B., Gehrke, J., & Knuf, L. (2000). Hierarchical coding in the perception and memory of spatial layouts. *Psychological Research*, 64, 1-10.
- Hommel, B., & Knuf, L. (2000). Action related determinants of spatial coding in perception and memory. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Spatial cognition II: Integrating abstract theories, empirical studies, formal methods, and practical applications* (pp. 387-398). Berlin: Springer.
- Hommel, B., Knuf, L., & Gehrke, J. (2002). Action-induced cognitive organization of spatial maps. Manuscript submitted for publication.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (in press). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24.
- Knuf, L., Klippel, A., Hommel, B. & Freksa, C. (2002). Perceptually induced distortions in cognitive maps. Manuscript submitted for publication.
- McNamara, T.P. (1986). Mental representation of spatial relations. *Cognitive Psychology*, 18, 87-121.
- McNamara, T.P. (1991). Memory's view of space. *Psychology of Learning and Motivation*, 27, 147-186.
- McNamara, T.P., Hardy, J.K., & Hirtle, S.C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 211-227.
- McNamara, T.P., & LeSueur, L.L. (1989). Mental representations of spatial and nonspatial relations. *Quarterly Journal of Experimental Psychology*, 41, 215-233.
- Palmer, S.E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9, 441-474.
- Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, 10, 422-427.
- Thorndyke, P. W. (1981). Distance estimation from cognitive maps. *Cognitive Psychology*, 13, 526-550.
- Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13, 407-433.
- Tversky, B., & Schiano, D.J. (1989). Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General*, 118, 387-398.

How Are the Locations of Objects in the Environment Represented in Memory?

Timothy P. McNamara¹

Department of Psychology, Vanderbilt University, 111 21st Ave South
Nashville, TN 37203
t.mcnamara@vanderbilt.edu

Abstract. This chapter summarizes a new theory of spatial memory. According to the theory, when people learn the locations of objects in a new environment, they interpret the spatial structure of that environment in terms of a spatial reference system. Our current conjecture is that a reference system intrinsic to the collection of objects is used. Intrinsic axes or directions are selected using egocentric (e.g., viewing perspective) and environmental (e.g., walls of the surrounding room) cues. The dominant cue is egocentric experience. The reference system selected at the first view is typically not updated with additional views or observer movement. However, if the first view is misaligned but a subsequent view is aligned with natural and salient axes in the environment, a new reference system is selected and the layout is reinterpreted in terms of this new reference system. The chapter also reviews evidence on the orientation dependence of spatial memories and recent results indicating that two representations may be formed when people learn a new environment; one preserves interobject spatial relations and the other comprises visual memories of experienced views.

1 Introduction

As any student of spatial cognition or geography knows, the concept of location is inherently relative. One cannot describe or specify the location of an object without providing, at least implicitly, a frame of reference. For example, the location of a chair in a classroom can be specified in terms of the room itself (e.g., the chair is in the corner by the door), other chairs in the room (e.g., the chair is in the first row, second column), or an observer (e.g., the chair is in front of me). Likewise, human memory systems must use spatial reference systems of some kind to preserve the remembered locations of objects.

There are many ways to classify spatial reference systems (Levinson, 1996), but a useful one, for the purposes of understanding human spatial memory, divides them

¹ Preparation of this chapter and the research reported in it were supported in part by National Institute of Mental Health Grant R01-MH57868. The chapter was improved as a result of the comments of two anonymous reviewers. I am enormously indebted to Vaibhav Diwadkar, Weimin Mou, Björn Rump, Amy Shelton, Christine Valiquette, and Steffen Werner for their contributions to the empirical and theoretical developments summarized in this chapter.

into two categories: *Egocentric reference systems* specify location and orientation with respect to the organism, and include eye, head, and body coordinates. *Environmental reference systems* specify location and orientation with respect to elements and features of the environment, such as the perceived direction of gravity, landmarks, or the floor, ceiling, and walls of a room.

The initial investigations of spatial reference systems conducted in our laboratory indicated that spatial memories might be defined egocentrically (e.g., Diwadkar & McNamara, 1997; Roskos-Ewoldsen, McNamara, Shelton, & Carr, 1998; Shelton & McNamara, 1997). For example, Shelton and McNamara (1997) required participants to learn the locations of seven objects in a room from two orthogonal viewpoints. After they had memorized the locations of the objects, the observers were escorted to a different room, on a different floor of the building, and asked to make judgments of relative direction using their memories (e.g., "Imagine you are standing at the shoe and facing the clock. Point to the jar."). These judgments were made with a computer mouse on a simulated dial and pointer displayed on the computer screen. Pointing judgments were faster and more accurate for imagined headings parallel to one of the two study views than for headings parallel to unfamiliar views. These results suggested that participants had formed two egocentric representations of the layout, one from each viewing position. We conceived of these representations as visual-spatial "snapshots" of the layout.

The results of subsequent investigations indicated that this conclusion was premature. Werner and Schmidt (1999) asked student residents of Göttingen, Germany to imagine themselves at the intersection of two major streets in town, facing in various directions, and then to identify landmarks in cued directions. They found that landmarks were identified faster and more accurately when the imagined heading was parallel to one of the major streets than when it was not (see also, Montello, 1991). This finding indicates that the students had represented the layout of the city in terms of reference axes established by the road grid.

More problematic still are results of experiments reported by Shelton and McNamara (2001b). In Shelton and McNamara's third experiment, participants learned the locations of objects in a room from two stationary points of view. One viewing position was aligned (0°) and the other was misaligned (135°) with a mat on the floor and the walls of the room (see Figure 1). Performance in subsequent judgments of relative direction indicated that the aligned view was represented in memory but the misaligned view was not (see Figure 2). Note that angular error in pointing judgments was as high for the familiar heading of 135° as for unfamiliar headings, even for participants who learned the view from 135° first! In another experiment, participants learned similar layouts in a cylindrical room from three points of view (0° , 90° , & 225°). Half of the participants learned the views in the order 0° - 90° - 225° , and half learned the views in the reverse order. Accuracy of judgments of relative direction indicated that only the first study view (0° or 225°) was mentally represented: Pointing judgments were quite accurate for imagined headings parallel to the first study view (mean error of 14.6°) but no more accurate for the second and third study views than for novel headings (mean error of 38.7° vs. 35.7° , respectively).

The visual-spatial snapshot model proposed by Shelton and McNamara (1997) would predict better performance on familiar than on unfamiliar headings. For example, in the cylindrical room experiment, it predicts, *ceteris paribus*, equally good performance on the headings of 0° , 90° , and 225° . The results of Werner and

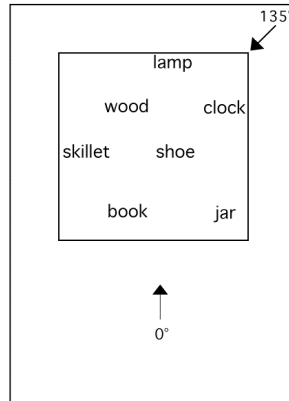


Fig. 1. Schematic illustration of one of the layouts used in Shelton and McNamara's (2001b) Experiment 3. Real objects were used, not names.

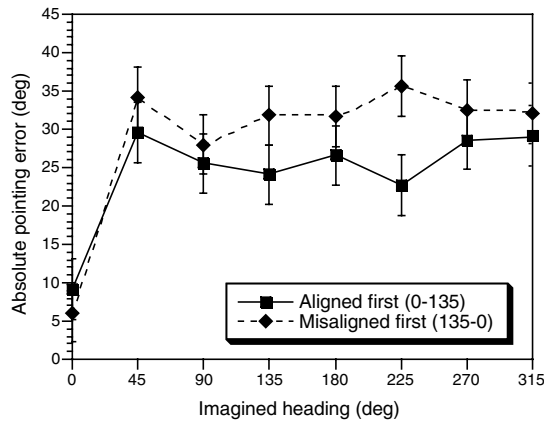


Fig. 2. Angular error in judgments of relative direction as a function of imagined heading and the order in which views were learned in Shelton and McNamara's (2001b) Experiment 3. Subjects learned an aligned view (0°) and a misaligned view (135°) of layouts similar to the one illustrated in Figure 1. Error bars are confidence intervals corresponding to ± 1 SEM as estimated from the ANOVA.

Schmidt's (1999) and Shelton and McNamara's (2001b) experiments indicated that spatial memories were not egocentric, and led to the development of the theory of spatial memory described in the next section.

2 Sketch of a Theory of Human Spatial Memory

The theory of spatial memory that we have developed to explain these findings is firmly rooted in principles of form perception proposed by Rock (1973). Rock wanted

to know why the perceived shape of a figure depends on its orientation. A square, for example, is seen as a square when an edge is on top, but is seen as a diamond when a vertex is on top. Rock was particularly interested in whether a change in orientation with respect to the observer or a change in orientation with respect to the environment was the principal cause of changes in perceived shape.

Rock's investigations indicated that for unfamiliar figures, changing egocentric orientation had little effect on perceived shape. However, when the orientation of a figure with respect to the environment was changed, the figure was seen as different and often not recognized at all. For example, Rock (1956) designed ambiguous figures so that they had different interpretations in different orientations; for instance, in one orientation, one of the figures looked like the profile of an old man, but when rotated 90 degrees, it looked like an outline of the U.S. The figures were presented to observers whose heads were tilted 90 degrees. When shown these ambiguous figures with heads tilted, observers typically reported seeing the environmentally upright figure rather than the retinally upright figure. Another way to describe these findings is that observers saw the shape defined by the environmental frame of reference rather than the shape defined by the egocentric frame of reference; indeed, they ignored the egocentric information to interpret the figure in terms of the environmental information.

Rock (1973) concluded that the interpretation of a figure depends on which part or region is assigned "top," and that a change in the assignment of this direction profoundly affects perceived shape. The top of a figure is normally assigned on the basis of the information provided by gravity or the visual frame of reference. Other sources of information can also be used, including egocentric orientation, instructions, intrinsic properties of the figure, and familiarity, but these sources were, according to Rock, typically less salient than environmental sources. More recent investigations (e.g., Friedman & Hall, 1996; McMullen & Jolicoeur, 1990) have shown that Rock might have underestimated the importance of retinal orientation in the perception of form. Even so, the general principle--that the perception of form involves the assignment of directions based on a spatial reference system--is sound.

According to our theory (Mou & McNamara, 2002; Shelton & McNamara, 2001b; Werner & Schmidt, 1999), learning the spatial structure of a new environment involves interpreting it in terms of a spatial reference system. This process is analogous to determining the top of a figure or an object; in effect, conceptual "north" is assigned to the layout, creating privileged directions in the environment (conceptual "north" need not, and usually will not, correspond to true or magnetic north or any other cardinal direction). Our working hypothesis is that the spatial structure of the environment is represented in terms of an intrinsic reference system (Palmer, 1989); one defined by the layout itself (e.g., the rows and columns formed by chairs in a classroom). Intrinsic directions or axes are selected using cues, such as viewing perspective and other experiences (e.g., instructions), properties of the objects (e.g., they may be grouped together based on similarity or proximity), and the structure of the environment (e.g., geographical slant). An important difference between form perception and spatial memory is that whereas figures in the frontal plane are oriented in a space with a powerful reference axis, namely, gravity, the locations of objects are typically defined in the ground plane, which does not have privileged axes or directions (e.g., humans cannot perceive magnetic fields). We therefore propose that the dominant cue in spatial memory is egocentric experience. The spatial layouts learned by participants in most of our experiments were composed of small, moveable

objects. In general, however, a spatial layout could be composed of large or stationary objects, such as mountain peaks, trees, buildings, doors, windows, and so forth. We would still expect in such cases for intrinsic directions or axes to be identifiable, and for some to be more salient than others.

The theory is perhaps best understood in the context of concrete examples. Consider, first, the cylindrical room experiment conducted by Shelton and McNamara (2001b, Exp. 7). According to the theory, when observers studied the layout from the first viewing position, they interpreted its spatial structure in terms of an intrinsic reference system aligned with their viewing perspective. When participants were taken to the second and third points of view, they continued to interpret the spatial structure of the layout in terms of the reference system selected at the first point of view, just as if they were viewing a (now) familiar object at novel orientations. This reference system remained the dominant one, even when participants were moved to the next two points of view, because the layout did not have salient alternative axes, and because no other point of view was aligned with a salient axis in the environment.

As another example, consider the experiment summarized above in which participants learned an aligned and a misaligned view of a layout of objects (Figures 1 & 2). Participants who first learned the aligned viewpoint (0°) represented the layout in terms of an intrinsic reference system aligned with their viewing perspective, the edges of the mat, and the walls of the room. When they moved to the misaligned viewpoint (135°), they still interpreted the layout in terms of the reference system established by the first, aligned view. Hence, performance in judgments of relative direction was best for the heading parallel to the aligned view, and was no better for the heading parallel to the misaligned view than for novel headings. Observers who first learned the misaligned view (135°) also must have interpreted the space in terms of a reference system defined by that view. This conclusion follows from the results of another experiment in which participants learned the same layout but only from the misaligned point of view (Shelton & McNamara, 2001b, Exp. 2). The results of this experiment showed that participants represented the layout from this single familiar view. Our hypothesis is that when participants were taken to the second, aligned viewpoint, they *re-interpreted* the spatial structure of the layout in terms of a reference system defined by the aligned view because it was aligned with salient axes in the environment (e.g., the edges of the mat and the walls of the room) and with egocentric experience (albeit, a new experience). After moving from a misaligned study view to an aligned study view, observers changed the definition of "north." A new spatial reference system—one that was aligned with the environment and egocentric experience—was selected and the spatial layout was reinterpreted in terms of it.

Our conjecture that spatial memories are defined in terms of intrinsic reference systems is supported by findings reported by Mou and McNamara (2002). They required participants to learn layouts like the one illustrated in Figure 3. Objects were placed on a square mat oriented with the walls of the enclosing room or on the bare floor of a cylindrical room. In one experiment, participants studied the layout from 315° , and were instructed to learn the layout along the egocentric 315° axis or the nonegocentric 0° axis. This instructional manipulation was accomplished by pointing out that the layout could be seen in "columns" consistent with the appropriate axis (e.g., *clock-jar*, *scissors-shoe*, etc. vs. *scissors-clock*, *wood-shoe-jar*, etc.), and by asking participants to point to the objects in the appropriate order when they were quizzed during the learning phase. All participants viewed the layout from 315° . After

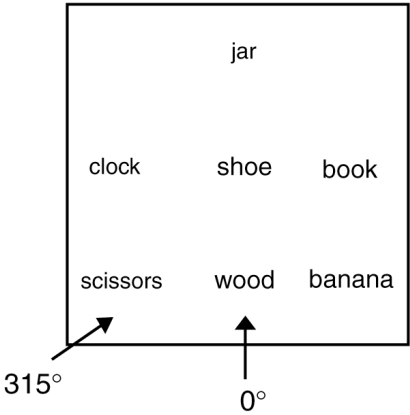


Fig. 3. Schematic illustration of one of the layouts used by Mou and McNamara (2002). Real objects were used, not names.

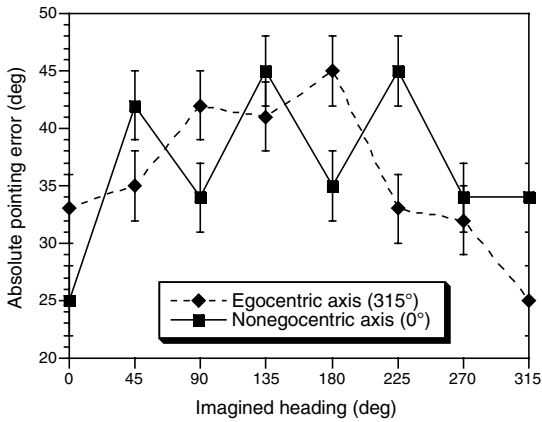


Fig. 4. Angular error in judgments of relative direction as a function of imagined heading and learning axis in Mou and McNamara's (2002) Experiment 2. All subjects viewed the layout in Figure 3 from 315°. They were instructed to learn the layout along the egocentric 315°-135° axis or the nonegocentric 0°-180° axis. Error bars are confidence intervals corresponding to ± 1 SEM as estimated from the ANOVA.

learning, participants made judgments of relative direction using their memory of the layout.

One important result (see Figure 4) is the near perfect crossover interaction for imagined headings of 0° and 315°: Participants who were instructed to learn the layout along the egocentric 315° axis were better able to imagine the spatial structure of the layout from the 315° heading than from the 0° heading, whereas the opposite pattern was obtained for participants who learned the layout along the nonegocentric 0° axis. In particular, participants in the 0° group were better able to imagine the spatial structure of the layout from an unfamiliar heading (0°) than from the heading they actually experienced (315°). A second important finding is the different patterns

of results for the two groups: In the 0° group, performance was better on novel headings orthogonal or opposite to 0° (90° , 180° , & 270°) than on other novel headings, producing a sawtooth pattern, whereas in the 315° group performance on novel headings depended primarily on the angular distance to the familiar heading of 315° . The sawtooth pattern in the 0° group also appeared when the objects were placed on the bare floor of a cylindrical room, which indicates that this pattern was produced by the intrinsic structure of the layout, not by the mat or the walls of the enclosing room. The third major finding was that there was no apparent cost to learning the layout from a nonegocentric perspective. Overall error in pointing did not differ across the two groups.

We believe that the sawtooth pattern arises when participants are able to represent the layout along two intrinsic axes (e.g., 0° - 180° and 90° - 270°). Performance may be better on the imagined heading of 0° because this heading was emphasized during the learning phase. We suspect that the sawtooth pattern did not occur in the condition in which participants learned the layout according to the 315° - 135° axis because the 45° - 225° axis is much less salient in the collection of objects. Indeed, we suspect that participants did not usually recognize that the layout could be organized along "diagonal" axes unless they actually experienced them because the "major" axes were much more salient; for example, the layout is bilaterally symmetric around 0° - 180° but not around 315° - 135° or 45° - 225° .

3 Alternative Theories

Aspects of our theoretical framework have been proposed or anticipated by others. Most notably, Tversky (1981) demonstrated that errors in memory of spatial relations could be explained in terms of heuristics derived from principles of perceptual organization, and argued that spatial memory was influenced by how a map or environment was interpreted when it was learned. She also discussed how intrinsic reference systems might be induced from features of the environment and used to represent location and orientation. Several experiments have demonstrated that spatial representations are influenced by the intrinsic structure of a layout or by the geometry of the surrounding environment (e.g., Easton & Sholl, 1995; Hermer & Spelke, 1994; Learmonth, Newcombe, & Huttenlocher, 2001; Montello, 1991; Werner & Schmidt, 1999).

Several influential models have been proposed to explain memory for location. Huttenlocher, Hedges, and Duncan (1991) and Lansdale (1998) have proposed elegant mathematical models of positional uncertainty and bias in memory of the location of single object. Neither of these projects was aimed at investigating the spatial reference systems used in memory, although Huttenlocher et al. concluded from the distributions of memory reports that participants used polar coordinates to represent the location of a single dot in a circle. It is not clear how these models could explain the orientation dependence of spatial memories, or how they could be scaled up to large-scale spaces.

The spatial-framework model investigated by Bryant, Franklin, and Tversky (e.g., Bryant & Tversky, 1999; Franklin & Tversky, 1990) is more relevant to the situations examined in our studies. In particular, Bryant and Tversky (1999) had participants study two-dimensional (2D) diagrams or three-dimensional (3D) models of six

objects surrounding a central character in the canonical directions front, back, right, left, head (e.g., above an upright character) and feet (e.g., below an upright character). In the test phase, the participants identified the objects in cued directions. Across trials, the central character was described as rotating to face different objects, and as changing orientation (e.g., from upright to reclining). Bryant and Tversky concluded that diagrams, and other 2D interpretations of the scenes, were represented using an intrinsic reference system centered on the character, whereas the models, and other 3D interpretations of the scenes, were represented with an egocentric spatial framework in which participants mentally adopted the orientation and the facing direction of the central character.

The use of an intrinsic reference system for 2D scenes is broadly consistent with our theoretical framework. As Bryant and Tversky (1999) use the term, it refers to an object-based reference system centered on objects that have intrinsic asymmetries, such as people and cars. In our theoretical framework, it refers to a reference system in which reference directions or axes are induced from the layout of the environment to be learned. The basic idea is similar, however. The egocentric spatial framework used for 3D scenes would seem to be inconsistent with our model. In fact, we believe the two are complementary. Bryant and Tversky's experiments examine situations in which the observer has adopted an orientation in imagination, and then is asked to retrieve objects in cued directions. The difficulty of retrieving or inferring the spatial structure of the layout from novel versus familiar orientations is not measured. Our experiments, in contrast, have focused on effects of orientation, not on the efficiency of retrieval of objects in cued directions. The results of experiments in which both effects have been assessed (e.g., Sholl, 1987; Werner & Schmidt, 1999) indicate that they may be independent.

The independence of egocentric and allocentric coding of spatial relations is embodied in Sholl's model of spatial representation and retrieval (e.g., Easton & Sholl, 1995; Sholl & Nolin, 1997). This model contains two subsystems: The *self-reference system* codes self-to-object spatial relations in body-centered coordinates, using the body axes of front-back, right-left, and up-down (as in the spatial framework model). This system provides a framework for spatially directed motor activity, such as walking, reaching, and grasping. The *object-to-object system* codes the spatial relations among objects in environmental coordinates. Spatial relations in this system are specified only with respect to other objects (i.e., an intrinsic reference system is used). Relative direction is preserved locally, among the set of objects, but not with respect to the surrounding environment, and there is no preferred direction or axis. The representation is therefore orientation-independent. These two systems interact in several ways. For example, the heading of the self-reference system fixes the orientation of the object-to-object system, in that the front pole of the front-back axis determines "forward" in the object-to-object system. As the self-reference system changes heading, by way of actual or imagined rotations of the body, the orientation of the object-to-object system changes as well.

At present, our theoretical framework does not address self-to-object spatial relations, although we recognize that such spatial relations must be represented, at least at the perceptual level, for the purpose of guiding action in space and seem to play an important role in the spatial-framework paradigm. An important similarity between Sholl's model and ours is the use of intrinsic reference systems to represent interobject spatial relations. A major difference, though, is that the object-to-object system is orientation independent in Sholl's model but orientation dependent in ours.

4 Orientation Dependence vs. Independence

Over the past two decades, a large number of experiments have examined, at least indirectly, the orientation dependence of spatial memories. Participants have learned several views of layouts; have learned layouts visually, tactilely, via navigation, and via desktop virtual reality; have been tested in the same room in which they learned the layout or in a different room; have been oriented or disoriented at the time of testing; have been seated or standing during learning and testing; and have been tested using scene recognition, judgments of relative direction, or both (e.g., Christou & Bühlhoff, 1999; Diwadkar & McNamara, 1997; Easton & Sholl, 1995; Levine, Jankovic, & Palij, 1982; Mou & McNamara, 2002; Presson & Montello, 1994; Richardson, Montello, & Hegarty, 1999, map & virtual-walk conditions; Rieser, 1989; Rieser, Guth, & Hill, 1986; Roskos-Ewoldsen et al., 1998; Shelton & McNamara, 1997, 2001a, 2001b, 2001c; Sholl & Nolin, 1997, Exps. 1, 2, & 5; Simons & Wang, 1998). A consistent finding has been that performance is orientation dependent. In most of those studies, orientation dependence took the form of better performance on familiar views and orientations than on unfamiliar views and orientations; in Mou and McNamara's (2002) experiments, performance was better on orientations aligned with the intrinsic axis of learning than on other orientations.

Orientation independent performance has been observed, however, in several published studies (Evans & Pezdek, 1980; Presson, DeLange, & Hazelrigg, 1989; Presson & Hazelrigg, 1984; Richardson et al., 1999, real-walk condition; Sholl & Nolin, 1997, Exps. 3 & 4). In a now classical study, Evans and Pezdek (1980) reported evidence of orientation independence in memory of a large-scale environment. Participants were shown sets of three building names, which were selected from the Cal State-San Bernardino campus, and had to decide whether or not the buildings were arranged in the correct spatial configuration. Incorrect triads were mirror images of correct triads. Participants in one experiment were students at the university who presumably learned the locations of buildings naturally via navigation; participants in another experiment were students at another university who had memorized a map of the Cal State-San Bernardino campus. The independent variable was the angular rotation of the test stimulus relative to the canonical vertical defined by the map. For students who had learned the map, the familiar upright views of the stimuli were recognized fastest, and the difficulty of recognizing unfamiliar, rotated stimuli was a linear function of angular rotation (e.g., Shepard & Metzler, 1971). However, for students who had learned the campus naturally, there was no such relation: Response times were roughly the same at all angles of rotation. An analysis of individual participants' data revealed no linear trends even when alternative canonical orientations were considered.

To our knowledge, Evans and Pezdek's (1980) experiments have never been replicated. One explanation for the pattern of results is that students who learned the campus experienced it from many points of view and orientations, whereas students who learned the map only experienced the map in one orientation. Recent evidence indicates, however, that learning a large-scale environment from several orientations is not sufficient to produce an orientation independent representation. McNamara, Rump, and Werner (in press) had student participants learn the locations of eight objects in an unfamiliar city park by walking through the park on one of two prescribed paths, which encircled a large rectangular building (a full-scale replica the

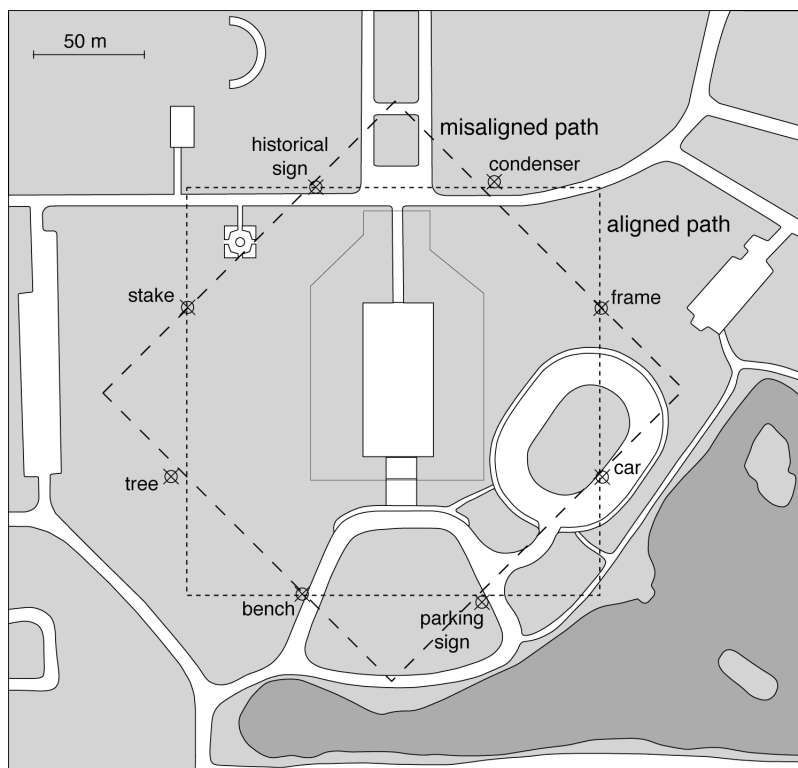


Fig. 5. Map of the park and paths in McNamara, Rump, and Werner's (in press) experiment. The white rectangle in the center is the Parthenon. Dark shaded area in lower right is the lake.

Parthenon in Athens, Greece). The aligned path was oriented with the building; the misaligned path was rotated by 45° (see Figure 5). Participants walked the path twice, and spent about 30 minutes learning the locations of the objects. They were then driven back to the laboratory, and made judgments of relative direction using their memories. As shown in Figure 6, pointing accuracy was higher in the aligned than in the misaligned path group, and the patterns of results differed: In the aligned condition, accuracy was relatively high for imagined headings parallel to legs of the path (0° , 90° , 180° , 270°) and for an imagined heading oriented toward a nearby lake, a salient landmark (225°). In the misaligned condition, pointing accuracy was highest for the imagined heading oriented toward the lake (a heading that was familiar), and decreased monotonically with angular distance. For both groups, though, performance was orientation dependent; there was no evidence that participants were able to construct view-invariant representations of the spatial structure of the park after experiencing it from four orientations.

In another influential line of research, Presson and his colleagues (Presson et al., 1989; Presson & Hazelrigg, 1984) obtained evidence that orientation dependence was modulated by layout size. Participants learned 4-point paths from a single perspective. These paths were small (e.g., 40 cm X 40 cm) or large (e.g., 4 m X 4 m). After

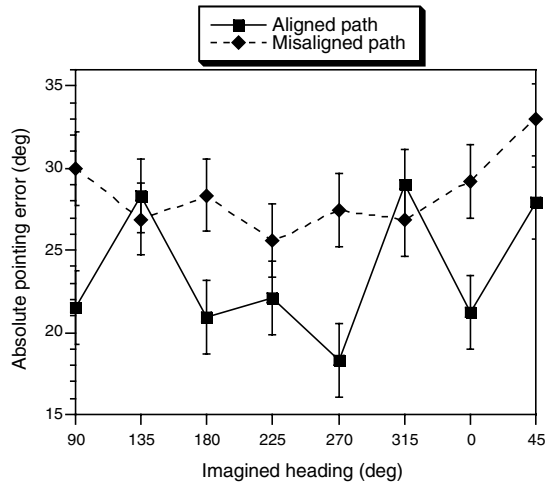


Fig. 6. Angular error in judgments of relative direction as a function of imagined heading and path. Subjects learned the locations of 8 objects in the park by walking either the aligned path or the misaligned path (see Figure 5). Data are plotted to emphasize the symmetry around the heading of 225°. Error bars are confidence intervals corresponding to ± 1 SEM as estimated from the ANOVA.

learning a layout, participants made judgments of relative direction using their memories of the layout. Imagined headings were aligned or contra-aligned with the original viewing perspective. The experiments showed that layout size was the only consistent predictor of the relative difficulty of aligned and contra-aligned judgments. When participants learned small layouts, aligned judgments were more accurate than contra-aligned judgments, but when they learned large layouts, the difference in accuracy was reduced or eliminated. This interaction occurred even though participants viewed small and large layouts from a single perspective.

Roskos-Ewoldsen et al. (1998) attempted to replicate the learning and the test conditions used by Presson and his colleagues (Presson et al., 1989; Presson & Hazelrigg, 1984), and yet still obtained orientation dependent performance. We learned after conducting the experiments that there were important differences in how participants were tested. Roskos-Ewoldsen et al. discussed these differences at length, and concluded that participants in Presson's experiments might have been able to update their mental representations when they were tested (e.g., Rieser, 1989).

Sholl and Nolin (1997) also attempted to replicate Presson, DeLange, and Hazelrigg's (1989) findings, and for the most part, were unable to do so. However, Sholl and Nolin were able to obtain orientation independent performance in one combination of conditions, namely, when participants learned the 4-point paths from a low viewing angle (e.g., while seated) and were tested in a condition in which their physical location and facing direction at the time they made their pointing judgment matched those specified in the judgment of relative direction. Unfortunately, these same learning and test conditions produced orientation dependent performance in experiments conducted by Mou and McNamara (2001), although their participants learned more complex layouts (seven objects distributed on the floor of a large room as opposed to 4-point paths).

Finally, Richardson, Montello, and Hegarty (1999) had participants learn the interior hallways of a large building by walking through the building, by navigating a desktop virtual environment, or by learning a map. Afterwards, participants engaged in several tasks, including pointing to target locations from imagined and actual locations in the building. Orientation dependence was tested in the virtual-walk and in the real-walk conditions by comparing pointing judgments for headings aligned with the first leg of the path to pointing judgments for other headings. Aligned judgments were more accurate than misaligned judgments in the virtual-walk condition but these judgments did not differ in the real-walk condition, suggesting that real movement in the space allowed participants to form orientation independent mental representations. It is possible that if alignment were defined with respect to a different reference axis (e.g., the longest leg of the path), or different reference axes for different participants, evidence of orientation dependence might appear (e.g., Valiquette, McNamara, & Smith, 2002).

An important feature of all of the experiments in which orientation independent performance has been observed, with the exception of the Evans and Pezdek (1980) experiments, is that only two orientation conditions were compared: In the aligned condition, the imagined heading was parallel to the learning view (e.g., in Figure 1, "Imagine you are at the book, facing the wood; point to the clock"), and in the contra-aligned condition, the imagined heading differed by 180° from the learning view (e.g., "Imagine you are at the wood, facing the book; point to the clock"). This fact may be important because performance in judgments of relative direction for the imagined heading of 180° is often much better than performance for other novel headings, and can be nearly as good as that for the learning view (see, e.g., Figure 4). The cause of this effect is not clear, but it is possible that, for as yet unknown reasons, participants sometimes represent, at least partially, the spatial structure of the layout in the contra-aligned direction. It is also possible that participants are able to capitalize on self-similarity under rotations of 180° under certain conditions (e.g., Vetter, Poggio, & Bülthoff, 1994). In our opinion, investigations of the orientation dependence of spatial memories are at a distinct disadvantage if only aligned and contra-aligned conditions are compared.

In summary, there may be conditions in which people are able to form orientation independent spatial representations but these situations seem to be the exception rather than the rule; in addition, attempts to replicate some of these findings have not been successful. In our opinion, the balance of evidence indicates that spatial memories are orientation-dependent.

5 One Representation or Two?

Recent experiments conducted in our laboratory suggest that at least two independent representations may be formed when participants learn a spatial layout visually. One of these representations seems to preserve interobject spatial relations, and is used to make judgments of relative direction, whereas the other is a visual memory of the layout, and supports scene recognition.

One of these experiments was an investigation of spatial perspective taking (Shelton & McNamara, 2001a). One participant (the director) viewed a display of objects from a single perspective and described the display to a second participant (the matcher) from a perspective that differed by 0° , 45° , 90° , 135° , or 180° from the viewing perspective (e.g., Schober, 1993). The matcher's task was to reconstruct the layout from the director's description. The two were separated by a barrier that prevented visual contact. After they had finished, the *director's* memory for the spatial layout was tested using judgments of relative direction, old-new scene recognition (e.g., Diwadkar & McNamara, 1997), and map drawing. We were particularly interested in the effects of describing the layout from a nonegocentric perspective on the director's memory of the layout. Angular error in judgments of relative direction is reproduced in Figure 7. These results indicated that the described view was represented in memory at least as well as, and in two conditions (viz., disparities of 45° and 135°) better than, the visually perceived view. By contrast, the results from scene recognition (Figure 8) indicated that only the visually perceived view was represented in memory; the described view was recognized no faster and no more accurately than were novel views (which had been neither seen nor described). Scene recognition is a visual task, so it is not surprising that participants could not recognize the "unseen" view as fast as they could recognize the visually perceived view. It is, however, intriguing that the described view, which clearly showed a benefit in judgments of relative direction, appeared to be no different from a novel view during scene recognition.

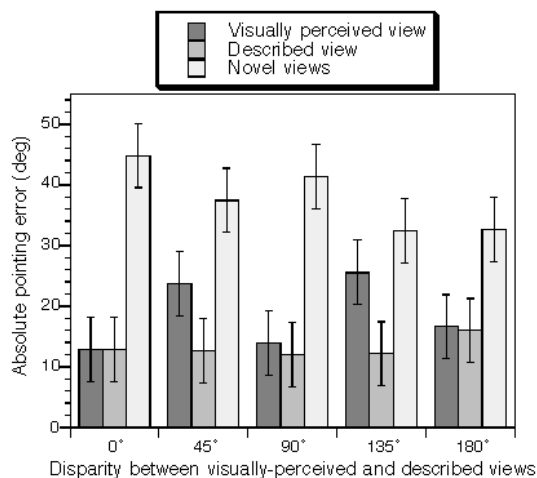


Fig 7. Angular error in judgments of relative direction in Shelton and McNamara's (2001a) experiment. Error bars are confidence intervals corresponding to ± 1 SEM as estimated from the ANOVA.

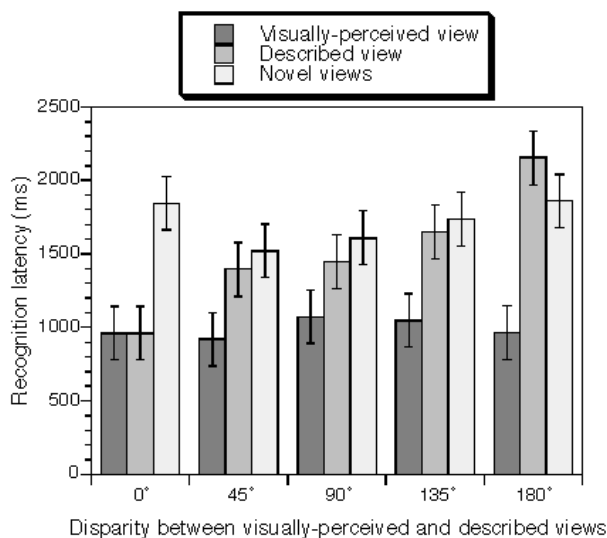


Fig. 8. Response latency in visual scene recognition in Shelton and McNamara's (2001a) experiment. Error bars are confidence intervals corresponding to ± 1 SEM as estimated from the ANOVA.

Additional evidence of multiple spatial representations can be found in an experiment just completed in our laboratory. The learning phase of this experiment replicated that of Shelton and McNamara's (2001b) Experiment 3, which was summarized earlier in this chapter (Figures 1 and 2). After learning, participants took part in two tasks: Old-new scene recognition, in which participants had to discriminate pictures of the layout, regardless of point of view, from pictures of the same objects in different spatial configurations; and judgments of relative direction. The results of judgments of relative direction (see Figure 9) largely replicated our original findings (see Figure 2), and indicated that the aligned view was mentally represented but the misaligned view was not: Performance for the familiar heading of 135° was worse than performance for the familiar heading of 0°, and not statistically better than performance for unfamiliar headings. As discussed earlier, we attribute the savings at headings of 90°, 180°, and 270° to partial representation of spatial relations along directions orthogonal and opposite to the primary intrinsic direction. The results from scene recognition (see Figure 10), however, showed that views of 0° and 135° were recognized equally well, and better than views from novel perspectives. This pattern indicated that both views were mentally represented. There were no effects of order of learning (0-135 vs. 135-0) in either task, and both graphs collapse across this variable.

6 Summary and Prospectus

Our primary goal in this chapter was to summarize a new theory of spatial memory. This theory, which is still in its infancy, attempts to explain how the locations of objects in the environment are represented in memory.

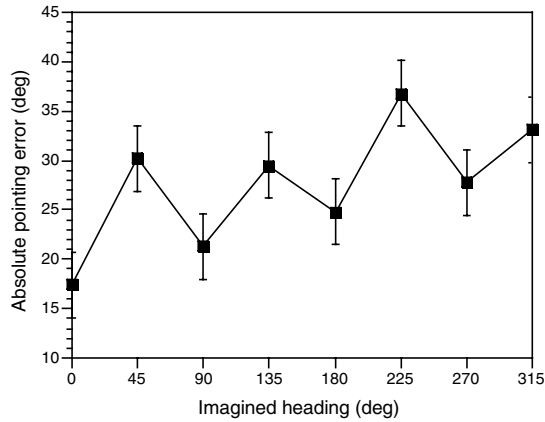


Fig. 9. Angular error in judgments of relative direction as a function of imagined heading. Subjects learned an aligned view (0°) and a misaligned view (135°) of layouts similar to the one illustrated in Figure 1. Error bars are confidence intervals corresponding to ± 1 SEM as estimated from the ANOVA.

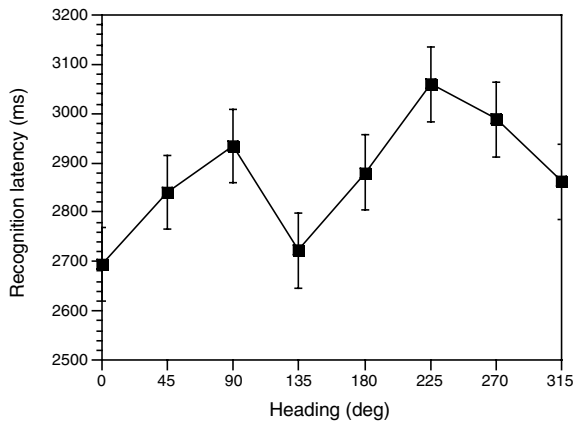


Fig. 10. Response latency in visual scene recognition as a function of heading. Subjects learned an aligned view (0°) and a misaligned view (135°) of layouts similar to the one illustrated in Figure 1. Error bars are confidence intervals corresponding to ± 1 SEM as estimated from the ANOVA.

According to the theory, when people learn a new environment, they represent the locations of objects in terms of a reference system intrinsic to the layout itself. Axes intrinsic to the collection of objects are selected and used to represent location and orientation. These axes are chosen on the basis of egocentric experience (including verbal instructions), spatial and nonspatial properties of the objects, and cues in the surrounding environment. We view this process as being analogous to identifying the top of a figure; in effect, conceptual "north" (and perhaps, east, west, & south) is created at the time of learning. Recent findings also suggest, however, that visual memories of familiar views are stored, regardless of their alignment with

environmental reference systems. The relationship between these two spatial representations is unknown at this time.

This theory makes the strong claim that spatial memories are composed primarily of object-to-object spatial relations, and therefore are allocentric. This claim conflicts with several recent proposals that spatial memories are primarily egocentric (e.g., Shelton & McNamara, 1997; Wang, 1999). Egocentric self-to-object spatial relations must be represented, at least at the perceptual level, for the purpose of guiding action in the environment (e.g., Anderson, 1999). It is an open question, however, whether egocentric spatial relations are represented in long-term memory. In principle, the spatial information needed for spatially directed motor activity could be computed from object-to-object spatial relations. Such a division of labor in spatial representation and processing between a transient egocentric system and a more permanent allocentric system bears strong resemblance to Milner and Goodale's (1995) account of dorsal and ventral streams of visual processing in the primate brain.

As people move through an environment, they must continuously update their location and orientation with respect to stable elements of the landscape to avoid getting lost or disoriented. Investigations of the ability to update with respect to a previously experienced collection of objects indicate that updating is of high fidelity and automatic, in the sense that it cannot be ignored (e.g., Farrell & Robertson, 1998; Rieser, 1989). Evidence from our laboratory indicates that the object-to-object system is not updated. For instance, if participants had updated in Shelton and McNamara's (2001b) cylindrical room experiment, one would expect performance to have been best on the heading parallel to the last study view, or perhaps on headings parallel to each of the three study views. In fact, spatial relations were retrieved efficiently only from the first study view. If the object-to-object system is not updated during locomotion, what is updated? Sholl and Nolin (1997) and Wang (1999) have suggested that egocentric self-to-object spatial relations are updated continuously as people move through an environment. It is also possible, however, that people update their position and orientation with respect to the same reference system used to represent the spatial structure of the environment, in effect treating their bodies as just another object in the space. Investigations of these questions should lead to a better understanding of how remembered spatial relations are used to guide action in space.

References

- Anderson, R. A. (1999). Multimodal integration for the representation of space in the posterior parietal cortex. In N. Burgess, K. J. Jeffery, & J. O'Keefe (Eds.), *The hippocampal and parietal foundations of spatial cognition* (pp. 90-103). Oxford: Oxford University Press.
- Bryant, D. J., & Tversky, B. (1999). Mental representations of perspective and spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 137-156.
- Christou, C. G., & Bühlhoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, 27, 996-1007.
- Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, 8, 302-307.
- Easton, R. D., & Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 483-500.

- Evans, G. W., & Pezdek, K. (1980). Cognitive mapping: Knowledge of real-world distance and location information. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 13-24.
- Farrell, M. J., & Robertson, I. H. (1998). Mental rotation and the automatic updating of body-centered spatial relationships. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 227-233.
- Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 63-76.
- Friedman, A., & Hall, D. L. (1996). The importance of being upright: Use of environmental and viewer-centered reference frames in shape discriminations of novel three-dimensional objects. *Memory & Cognition*, 24, 285-295.
- Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, 370, 57-59.
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, 98, 352-376.
- Lansdale, M. W. (1998). Modeling memory for absolute location. *Psychological Review*, 105, 351-378.
- Learmonth, A. E., Newcombe, N. S., & Huttenlocher, J. (2001). Toddlers' use of metric information and landmarks to reorient. *Journal of Experimental Child Psychology*, 80, 225-244.
- Levine, M., Jankovic, I. N., & Palij, M. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology: General*, 111, 157-175.
- Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 109-169). Cambridge, MA: MIT Press.
- McMullen, P. A., & Jolicoeur, P. (1990). The spatial frame of reference in object naming and discrimination of left-right reflections. *Memory & Cognition*, 18, 99-115.
- McNamara, T. P., Rump, B., & Werner, S. (in press). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonomic Bulletin & Review*.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Montello, D. R. (1991). Spatial orientation and the angularity of urban routes: A field study. *Environment and Behavior*, 23, 47-69.
- Mou, W., & McNamara, T. P. (2001). *Spatial memory and spatial updating*. Unpublished manuscript.
- Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 162-170.
- Palmer, S. E. (1989). Reference frames in the perception of shape and orientation. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 121-163). Hillsdale, NJ: Erlbaum.
- Presson, C. C., DeLange, N., & Hazelrigg, M. D. (1989). Orientation specificity in spatial memory: What makes a path different from a map of the path? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 887-897.
- Presson, C. C., & Hazelrigg, M. D. (1984). Building spatial representations through primary and secondary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 716-722.
- Presson, C. C., & Montello, D. R. (1994). Updating after rotational and translational body movements: Coordinate structure of perspective space. *Perception*, 23, 1447-1455.
- Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition*, 27, 741-750.
- Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1157-1165.

- Rieser, J. J., Guth, D. A., & Hill, E. W. (1986). Sensitivity to perspective structure while walking without vision. *Perception, 15*, 173-188.
- Rock, I. (1956). The orientation of forms on the retina and in the environment. *American Journal of Psychology, 69*, 513-528.
- Rock, I. (1973). *Orientation and form*. New York: Academic Press.
- Roskos-Ewoldsen, B., McNamara, T. P., Shelton, A. L., & Carr, W. (1998). Mental representations of large and small spatial layouts are orientation dependent. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 215-226.
- Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*, 1-24.
- Shelton, A. L., & McNamara, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review, 4*, 102-106.
- Shelton, A. L., & McNamara, T. P. (2001a). *Spatial memory and perspective taking*. Unpublished manuscript.
- Shelton, A. L., & McNamara, T. P. (2001b). Systems of spatial reference in human memory. *Cognitive Psychology, 43*, 274-310.
- Shelton, A. L., & McNamara, T. P. (2001c). Visual memories from nonvisual experiences. *Psychological Science, 12*, 343-347.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.
- Sholl, M. J. (1987). Cognitive maps as orienting schemata. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 615-628.
- Sholl, M. J., & Nolin, T. L. (1997). Orientation specificity in representations of place. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1494-1507.
- Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science, 9*, 315-320.
- Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology, 13*, 407-433.
- Valiquette, C. M., McNamara, T. P., & Smith, K. (2002). *Locomotion, incidental learning, and the orientation dependence of spatial memory*. Unpublished manuscript.
- Vetter, T., Poggio, T., & Bülthoff, H. H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology, 4*, 18-23.
- Wang, R. F. (1999). Representing a stable environment by egocentric updating and invariant representations. *Spatial Cognition and Computation, 1*, 431-445.
- Werner, S., & Schmidt, K. (1999). Environmental reference systems for large scale spaces. *Spatial Cognition and Computation, 1*, 447-473.

Priming in Spatial Memory: A Flow Model Approach¹

Karin Schweizer¹

¹University of Wuppertal, Gauß-Str. 20,
D-42097 Wuppertal, Germany
(kschweiz@t-online.de)

Abstract. Theories on spatial priming usually explain the reduction of reaction time (the priming effect) by spreading activation. In the field of spatial cognition concurrent models like post-lexical priming mechanisms or compound cue theories (expectancy-based priming theories) have not been consequently discussed. None of the existing theories, however, provides a sufficient explanation for both kinds of findings, various distance effects and alignment effects in episodic spatial memory. Moreover, all existing theories need a series of additional assumptions and transformations to translate theoretical magnitudes like activation or familiarity into reaction time latencies. the transformation from activation or familiarity to reaction times. This unsatisfying state of the art implies to suggest a new approach to think about spatial priming. The illustrated solution regards priming as a specific solution of the Navier-Stokes equation. Empirical data support the suggested model.

1 Introduction

In this paper I choose a flow model approach to describe priming processes. The hereby selected specific solution of the Navier-Stokes equation seems to be a reasonable solution because none of the existing theories is able to provide uniform explanations concerning results in (spatial) priming studies. On the one hand, no single theory can account for various semantic priming results like nonword facilitation, mediated priming, and backward priming for example (see Neely, 1991). On the other hand, existing theories need a series of additional assumptions and transformations to translate theoretical magnitudes like activation or familiarity into reaction time latencies (see below). This applies to spatial priming theories, too. To maintain existing theories, diverse effects on distance and direction (alignment) are explained with an overload of assumptions affecting the represented spatial memory. This theoretical level is unsatisfying:

The proposed flow model approach tries to improve priming theories. First of all, it allows to integrate findings in spatial priming and to match reaction time latencies to priming velocities. Secondly, the description of a priming process as a certain flow is not restricted to spatial priming but might also be referred to recall processes in general. To point out these benefits of a flow model approach, I start to explain priming mechanisms and try to illustrate assumptions on theories of memory and

¹ This work was partly supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) in the framework of the Spatial Cognition Priority Program (He 270/19-1).

transformation rules. In a further chapter I list main results which were especially found in spatial priming studies and discuss how automatic spreading activation has been used to provide general accounts of spatial priming effects. Chapter 4 outlines a model of the suggested flow and chapter 5 and 6 supply an empirical evaluation and the final discussion of this approach.

2 Priming Mechanisms and Representation of Memory

From the late 70ies until now priming in spatial memory has been discussed heavily; a fact which is reflected by a large number of articles (e.g., McNamara, 1986; McNamara, Ratcliff, & McKoon, 1984; Clayton & Habibi, 1991; Curiel & Radvansky, 1998; McNamara, Hardy, & Hirtle, 1989; Wagener & Wender, 1985). To some extent the influence of priming studies can be explained by the method which can be traced back to experiments by Beller (1971), Meyer and Schvaneveldt (1971) as well as to Posner and Mitchell (1967). The priming method offers the opportunity to investigate retrieval mechanisms beyond respondents' deliberate answers. The method² involves presenting a respondent with two stimuli, usually simultaneously or with a very short temporal delay. The stimulus which is presented (or processed) at first is called the prime; the stimulus which follows is called the target. The time between the presentation of the prime and the exposure to the target is called SOA (stimulus onset asynchrony). The presentation of a suitable prime obviously effects respondents' reaction times. The reaction times needed to recognize or categorize stimuli related to that prime are shortened compared to reaction times without (related) primes (= priming effect). So far, most of the researchers in priming would surely agree. Opinions about the underlying mechanisms and representation structures, however, differ a lot.

In generally, priming is explained by three diverse mechanisms: spreading-activation, expectancy-based priming (e.g., compound-cue theories), and post-lexical priming mechanisms. These mechanisms are furthermore linked to specific underlying theories of memory. Spreading activation theories, for example, conceive the representation of memory as a network of semantic or sublexical units (e.g., Anderson, 1983; 1993; Collins & Loftus, 1975; McClelland & Rumelhart, 1981). Networks consist of such units (nodes) and connections between them, which represent the relations between nodes. By activating the corresponding node information is retrieved. Related information is activated by the connections between nodes as activation is spreading (Anderson, 1983; McNamara, 1992a, b). This process is attention related. On the one hand, the corresponding node serves as a source of activation as long as the questioned information is focused. On the other hand, activation decays if the focus of attention changes.

Compound-cue theories, which are regarded as a specification of expectancy-based priming mechanisms, often consider the represented information as a matrix of associations (e.g., McKoon & Ratcliff, 1992; Ratcliff & McKoon, 1981; 1988). In this sense, memory consists of numerous traces, sometimes called pictures, containing specific items, relations between items or relations between items and the learning

² Like Neely (1991) I focus primarily on data collected in a single word priming paradigm, in which the prime to which no overt answer is required is followed by a target object.

context (e.g., Gillund & Shiffrin, 1984; Murdock, 1982; Raaijmakers & Shiffrin, 1981). Corresponding models regularly vary from sets of attributes to vector models. Nevertheless, the retrieval of information is explained uniquely (e.g., SAM). Presenting a cue activates all associated pictures (items, items-to-items or context-to-items relations). The strength of activation is determined by a so called familiarity index, which reflects the association of the presented cue with the pictures in memory.

In his overview about priming mechanisms Neely (1991, see also Neely & Keefe, 1989) concludes that none of the before enumerated theories is able to explain all existing priming effects. Therefore, a third type of mechanism was specified: post-lexical priming mechanisms like post-lexical semantic matching. According to this mechanism semantic similarities of target objects are compared post lexically. Since semantic similarities only occur when dealing with word targets decisions about words or non-words are easily made. Post lexical semantic matching between two words, however, is assumed to be very time expensive (see also Neely, 1991; De Groot, 1985). I therefore conclude that post lexical semantic matching processes are of subordinate significance to the present research paradigm and restrict the discussion to the two above mentioned mechanisms: spreading activation and compound-cue theories.

Besides the fact that none of the described mechanisms can account for a considerable number of existing priming effects, spreading activation theories and compound-cue models raise another problem. None of the theories provide an explanation for the reduction of reaction times without any transformation. Typically, activation is transformed into time. In spreading activation theories the transformation of activation is given as (Anderson, 1983):

$$RT = I + \frac{1}{A} - Ke^{\frac{KA}{(1-e^{-KA})}}, \quad (1)$$

whereas I means a general reaction time latency, A means the strength of activation (which is computed by the sum of all nodes multiplied with a weighted function), and K holds for the upper fence of the reaction time.

Compound cue models even forego transformation rules between familiarity and reaction times. The transformation is generally regarded as a diffusion process which can be considered as a continuous version of a time related random process (Ratcliff, 1978). A rule to compute familiarities into reaction times still lacks.

To summarize, priming theories make at least three kinds of assumptions. Firstly, there is an assumption about the structure of the represented memory (e.g., network vs. SAM), secondly, a specific process is supposed (e.g., spreading activation vs. compound-cue), and thirdly, the presumptions of transformation rules are not clearly verbalized (e.g., linear vs. exponential transformation). Altogether, it has to be realized that priming theories hold an unsatisfying status, which should encourage us to think about new approaches.

3 Results of Spatial Priming Studies

Studies on spatial priming investigate three main topics. First of all, it is important whether information is stored as pure spatial information or rather as temporal

information. In an influential study McNamara, Ratcliff, and McKoon (1984) analyzed whether the subjective distance between two towns on a map is merely determined by the Euclidean distance or also by the route distance. With the term route distance they referred to the distance between two towns via a route drawn on a map. They presented their respondents with maps of locations which could be connected through routes or not. In a successive priming phase reaction times between near and far locations were measured. It could be shown that reaction time latencies depended on both, the Euclidean and the route distance (see also McNamara & LeSuer 1989; Merrill & Baird, 1987; Schweizer, 1997).

On the other hand, Clayton and Habibi (1991; see also Curiel & Radvansky, 1998) argued that in the above illustrated study the Euclidean distance effect was confounded with the temporal presentation of locations. The authors were able to show that priming effects concerning Euclidean distances only occur if the corresponding spatial layout is (re)presented as a map. In their investigations both research teams aggravated the learning of spatial positions. They even encouraged their subjects to learn the spatial positions of certain objects listwise. Thus, they failed to show a general distance effect. Priming effects only occurred according to the order of presentation. A fact which leads to the assumption that distance effects in spatial priming are strongly related to the learning procedure.

A second related topic is the nature of spatial information, as to say, are there hints to assume that spatial information is metric. This discussion also concerns the question whether spatial knowledge is represented in hierarchies. To explain the above discussed distance effects, McNamara and colleagues (McNamara, 1986; 1991) assumed that spatial information might be coded in several ways and in form of hierarchies. According to this assumptions the encoding of a layout starts with the decomposition in main regions represented through an "image unit" and containing the basic spatial information like the distance between two towns. Here, the term "region" is important. Regions might be settled by physical borders or in a subjective way by projecting internal margins to a spatial layout. Regions are often defined by landmarks, eye-catching objects on the way or widely seen historical or characteristic buildings (Downs & Stea, 1973; Hardwick, Woolridge, & Rinalducci, 1983; Janzen, Herrmann, Katz, Schweizer, 2000; Kitchin, 1994; Kitchin & Friendschuh, 2000; Kuipers, 1978; 1983; Lynch, 1960; Pick, Montello, & Somerville, 1988; Schölkopf & Mallot, 1995; Steck & Mallot, 2000; Tversky, 1981; Werner, Krieg-Brückner, Mallot, Schweizer, & Freksa, 1997).

At the end of the encoding process, spatial relations are encoded as kind of objects, as part of regions, as regions, and as the whole layout. In addition, one part of the information is encoded as metric and another part as nonmetric information (McNamara, Halpin, & Hardy, 1992). Finally, all these assumptions lead to the following equation on spatial priming effects (McNamara, 1986):

$$A_i = C_i + M \sum_j (R_{ij} L_{ij} A_j), \quad (2)$$

whereas A means the strength of activation, i and j are certain objects of the layout (node i and node j) represented, R holds for the strength of relation between the nodes, L for the probability (or likelihood) of the relation (which is related to the distance

and another factor called alignment, and the strength of activation of the related node)³.

Again, the resulting activation magnitude must be transformed into reaction time latencies or priming effects, which are computed of reaction times differences. The fact that reaction time latencies are sometimes compared directly to familiarities does not release from this intermediate step. Then, the transformation is left to the reader.

In equation 2, a factor named alignment is mentioned. The term alignment refers to the direction between two objects considered from a spectator's point of view. This factor is the third main topic in spatial priming studies (e.g., McNamara, 1986; Schweizer & Janzen, 1996; Schweizer, Herrmann, Janzen, & Katz, 1998). Alignment studies can be traced back to the investigations of Levine, Jankovich, and Palij (1982). The authors presented their respondents with maps of spatial locations. The maps also contained a certain route which connected the locations. The respondents were instructed to learn the spatial layout. They afterwards took part in a pointing task. Less mistakes were found when the map before learned was aligned with the orientation in the pointing task. This alignment effect shows that map learning is orientation specific. Similar results were found by Presson, DeLange, and Hazelrigg (1987) or Roskos-Ewoldson, McNamara, Shelton, and Carr (1998).

The above mentioned and further experiments by May, Péruch, and Savoyant (1995), for example, showed that orientation specific learning is not due to map learning. All early stages of spatial representation seem to contain the information concerning the point of view (see Chown, Kaplan & Kortenkamp, 1995; Franklin, Tversky, & Coon, 1992; Schweizer et al., 1998; Siegel & White, 1975; Strohecker, 2000). Most researches, however, argue that the importance of orientation specific information decreases with increasing experience. To argue against this assumption, Schweizer et al. (1998; see also Schweizer, 1997; Schweizer & Janzen, 1996) conducted a series of experiments which have show an effect of orientation specific learning. In one of these experiments, respondents were given route knowledge via a film of a spatial layout. Several times respondents saw this film, which could show the layout from point A to point Z or from point Z to point A. Subsequently, they took part in a priming phase during which the prime-target combinations, which had previously been shown as figures on flags along the route from A to Z or Z to A, were presented. This prime-target pairs differed according to the distance and to the alignment with the experienced direction of the film (route direction). Both factors evoked a significant reduction of the reaction time latencies (Schweizer et al., 1998), which confirms that distance as well as alignment (here the route direction) are important information units in spatial cognition.

Existing priming theories like spreading activation or compound-cue models should not only be able to explain various distance effects but also such alignment (or route direction) effects. Superficially, the factor alignment mentioned with equation (2) provides this possibility. The probability of the relation between two nodes in this equation is related to the distance and the alignment between the locations which are represented by the corresponding nodes. The network referring to the spatial layout should therefore not only contain (weighted) connections according to kinds of objects, parts of regions, regions, and the whole layout but also to the type of alignment (or even the type of orientation). This latter relation, however, is not yet

³ *C* and *M* are specific magnitudes which refer to self excitation and maintenance of activation (McNamara, 1986).

specified in related theories of spatial memory. It is not outlined clearly whether spatial orientation is generally represented twice or not⁴. Again, an unsatisfying and overloaded status of priming theories is stated. Since I finally did not find any explanation concerning spatial priming effects and compound-cue models it seems to be indicated to think about new approaches.

4 A Flow Model of Spatial Priming

As known from field theory or hydrodynamics, a flow field can be described through the following physical quantities⁵: velocity ($\vec{v}(x,y,z)$), pressure (p), density (ρ), and temperature (T). Overall, there are six equations to determine these variables. To describe a flow model of spatial priming, however, it is necessary to explain two different things: the state space and the corresponding flow. The state space contains all parameters which are necessary to determine the corresponding system (see also Abraham & Shaw, 1985). To identify these parameters, it seems now necessary to describe the present problem in detail (see also Schweizer, 2001).

The process to be described is regarded as a retrieval process which starts as soon as someone remembers a certain spatial layout. This process is initiated by the perception and recognition of one of the objects of the mentioned layout (the prime object). The prime object accelerates the recognition or identification of a second associated object (the target object). The respondents react within a certain period of time⁶. These reaction time latencies can be combined with the relations between prime and target. If we understand the terms near and far literally, reaction times for certain close and far related pairs of objects can be computed as velocities. The computation of those velocities is the first step to a direct access concerning the modeling of a priming process. The next step is the assignment of the quantities.

I assume that the perceiving of a prime can be described as the starting of a flow. The process runs in a fluid which can be regarded as the mental representation of the spatial layout. The process contains a change in time. As known from hydrodynamics, changes in time are determined by a dynamic view. Changes in time are discussed by equations of motion. Besides the kinematic description, equations of motion also provide the description of viscosity or inertia and volume vectors as well as surface vectors. Unfortunately, it is often not possible to determine these quantities without specific constraints. In the present problem one of these constraints is the incompressibility of the fluid. Then the Navier-Stokes equation becomes:

⁴ First approaches that are conceived with route graphs are pointed out in Werner, Krieg-Brückner, and Herrmann (2000).

⁵ The following elaborations can be referred to Milne-Thomson (1976), Birkhoff (1978), and Zierep (1997).

⁶ The measured reaction times are results of the recognition of various objects. Besides the time for the decision whether an object is part of the layout or not (recognition task), they also comprise the time for identifying the prime, identifying the target, and preparing a motor reaction. Therefore, the assumptions I make are not valid for any times other than the whole reaction times (see also Luce, 1986).

$$\frac{d\vec{v}}{dt} = \vec{f} - \frac{1}{\rho} \text{grad}p + y\Delta\vec{v} . \quad (3)$$

Further constraints for specific solutions of the Navier-Stokes equation are given in a plane Couette flow which passes between two plates with a distance a , a plate which rests and a plate which moves with a certain velocity (U). In this case the velocity of the flow field is determined by a dimensionless gradient of pressure (P):

$$P = -\frac{a^2}{2\eta U} \frac{dp}{dx} , \quad (4)$$

where η stands for the viscosity of the fluid.

With these constraints it is possible to determine the velocity of the flow field by the following formulation:

$$v_y = \frac{y}{a} \left[U - \left(1 - \frac{y}{a} \right) \frac{a^2}{2\eta} \frac{dp}{dx} \right], \text{ for } \frac{dp}{dx} = \text{const.} , \quad (5)$$

For various magnitudes of P the velocity of the flow field shows different slopes. If $P = 0$, the slope corresponds to the plane Couette flow, a monotone linear increase of velocity from the resting to the moving plate. The slope, however, becomes non-linear as soon as P increases or decreases (see also figure 1).

To apply this solution to the present problem, the state space and the corresponding rheological model, which is illustrated with figure one, must be defined. Therefore, the following assumptions are suggested:

1. The presentation of a prime starts a process during which specific objects of a perceived spatial layout are remembered.
2. This priming processes can be considered as a flow between a resting and one or two moving plates.
3. The plates are situated in a certain distance. In the case of episodically remembered spatial layouts, this distance corresponds to the maximal remembered distance of the layout.
4. The distance between objects of the layout is given through the distance a .
5. This distance might differ depending on the alignment of the layout. Aligned relations might be remembered longer than misaligned relations. In this case, the flow process passes between two moving plates and one resting plate.
6. For the present problem, the velocity of the moving plates (U) is constant.
7. This is also true for the dynamic viscosity (η).
8. The pressure gradient according to the flow direction, dp/dx , is constant but different from zero.

Figure 1 illustrates the resulting rheological model of the flow process which can be regarded as model for priming processes.

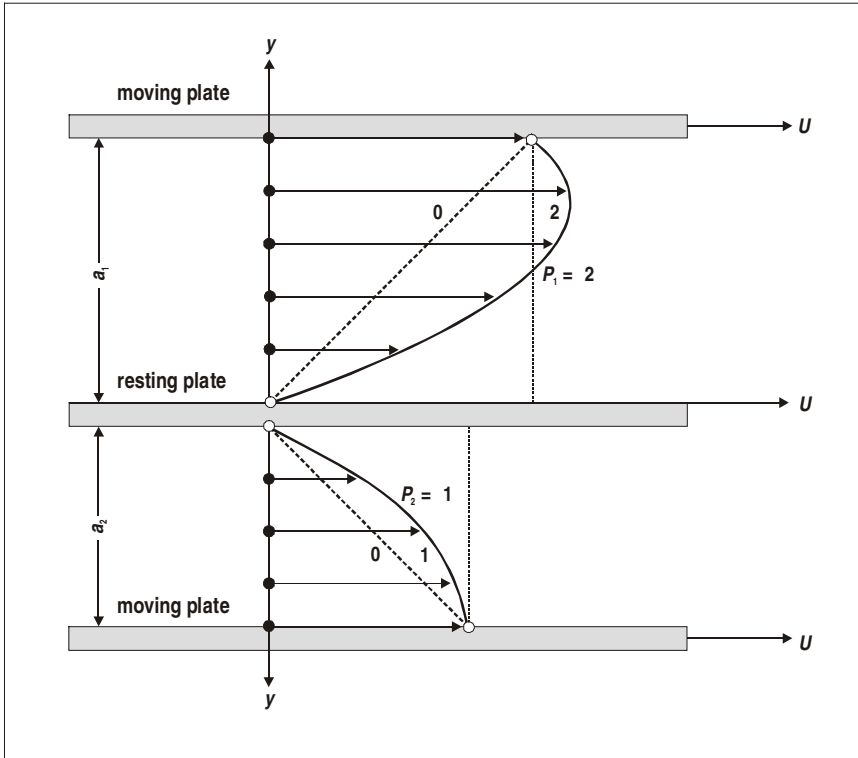


Fig. 1. Rheological model of the flow process

With these assumptions it is now possible to describe spatial priming as a process which evokes reaction time latencies that can be compared to velocities of a flow. As mentioned above, the referencing of reaction time latencies to physically measured distances enables the calculation of corresponding velocities. These velocities increase in a non-linear way if the gradient according to the flow direction (dp/dx) differs from zero. That means that reaction times change with the distance. In addition, the circumstance that distances might vary with the alignment of the representation provides different reaction time latencies for orientation specific prime-target pairs.

5 Experimental Evaluation of the Model

To evaluate the outlined model, data of a priming experiment carried out in a spatial priming study were re-analyzed (see Schweizer et al., 1998).

5.1 Experimental Procedure

In this experiment, the respondents took part in a similar priming task as described above. This time, a virtual environment including a film sequence (frame rate: 14 frames per second) was exposed. Figure 2 shows a plan of the spatial configuration. The virtual environment was based on a u-shaped space with 12 objects. The total layout appeared to have a length of 65,17 meters and a width of 25 meters in relation to the simulated eye height of the observer (1,70 meters). The objects were articles typically found in an office. They were standing on small pedestals. The room was presented as a museum for the office equipment of famous people. The film sequence started at "pot-plant" and led "clockwise" past the individual objects. The objects introduced in the film could be combined to make prime-target pairs which could be classified according to the distance and the direction of acquisition (alignment).

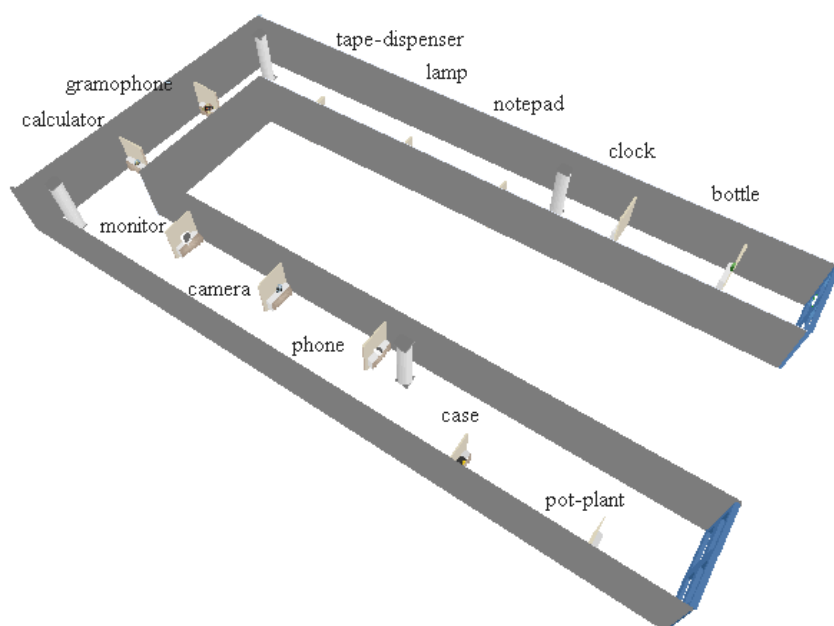


Fig. 2. Spatial Layout of the experiment described in the text

After having seen the films several times respondents were presented with prime-target pairs consisting of objects of the layout. The prime-target pairs were shown successively on a computer screen. The presentation time for the prime was 100 ms; the SOA was 350 ms. The target disappeared after the respondent had reacted. There was an interval of 1000 ms between respondents' reaction and the next prime. The images used as targets were either those which had been perceived before or unknown objects (distractor stimuli). As primes I only used images of objects which had already been seen. The respondents' task was to decide whether the presented image had been in the original scene or not (recognition task). The respondent had to press one of two keys for yes or no. Respondents' reaction times as well as the reactions

themselves (yes/no) were measured. The baselines (reaction time latencies for the same targets without primes) were measured in a separate procedure (Janzen, 2000).

5.2 Results

The recorded reaction times were corrected and averaged across the aligned and misaligned distances. Since I wanted to analyze the data with respect to their relations (near vs. far and aligned vs. misaligned), I firstly categorized the varying distances. Near prime-target pairs (items) were assigned to distances up to 11.6 meters; far items were assigned to distances from 25.5 to 33 meters in the model. The results of the computed ANOVA is illustrated in table 1.

Table 1. Strength of priming effects

Type	<i>M</i>	<i>SD</i>	Priming effects
All Items	700.96	92.94	Priming effect ¹ :
Baseline	768.65	92.62	-67.69**
Near	685.66	87.21	Priming effect ² :
Far	746.74	173.22	-61.13 ⁺
Aligned	671.80	102.40	Priming effect ² :
Misaligned	760.65	172.71	-88.86*

Legend: *M* = mean reaction time latencies in ms; *SD* = Standard deviation; ¹: difference between reaction time latencies with and without primes (baseline); ²: difference between items of the same relation, ⁺: statistical level at 10%; *: statistical level at 5%; **: statistical level at 1% or less.

A computed t-test shows an effect between primed reaction time latencies and the baseline ($t = 3.26$, $P < .005$). Furthermore, the afterwards computed ANOVA provides an effect ($F(1,19) = 6.41$, $P < .05$) concerning the alignment (route direction) of the objects in a spatial layout and also an important difference between near and far items ($F(1,19) = 3.88$, $P = .06$).

5.3 Modeling of the Priming Process

As mentioned above the first step to model the process component consists in calculating velocities for each one of the exposed prime-target pairs. Table 2 shows the results of the computed velocities for each one of the prime-target pairs.

The next step was to determine a velocity function for these empirical data. For this purpose, a regression function was estimated. This computation was carried out with respect to assumption 5 to 8. This means that P , in the case that dp/dx is constant but different from zero, has an influence on the computed velocities. P evokes non-linear slopes (see also fig. 1). Then, to all appearances the correlation should be quadratic. An accurate determination of the correlation, however, depends on the other quantities of equation (4) or equation (5). In the above described experiment, these quantities were constant except for the maximal remembered aligned or misaligned distances. Therefore, I chose two quadratic regressions to model the

empirical data, one for aligned (forward) and one for misaligned (backward) items (see equation 6 and 7).

$$v_{aligned} = 1,52d - 0,0033d^2 \quad , \quad (6)$$

$$v_{misaligned} = 1,45d - 0,0029d^2 \quad . \quad (7)$$

Table 2. Mean reaction time latencies for all prime-target pairs

Item	<i>d</i>	<i>M</i>	<i>V</i>
lamp – notepad	11.00	643.80	17.08
notepad – lamp	11.00	752.80	14.61
phone – camera	11.36	655.00	17.34
camera – phone	11.36	671.84	16.90
camera – monitor	11.52	643.90	17.89
monitor – camera	11.52	721.65	15.97
tape-dispenser – lamp	11.59	666.05	17.40
lamp – tape-dispenser	11.59	726.70	15.95
notepad – clock	11.60	653.00	17.76
clock – notepad	11.60	625.50	18.54
case – phone	11.86	696.50	17.02
phone – case	11.86	644.55	18.40
gramophon – tape-dispenser	12.07	832.06	14.50
tape-dispenser – gramophon	12.07	733.83	16.44
monitor – calculator	12.10	704.58	17.17
calculator – monitor	12.10	768.70	15.74
camera – notepad	18.42	620.89	29.67
notepad – camera	18.42	688.65	26.75
phone – lamp	18.64	716.26	26.02
lamp – phone	18.64	788.77	23.63
phone – clock	20.89	649.55	32.16
clock – phone	20.89	737.91	28.31
case – notepad	21.91	684.40	32.02
notepad – case	21.91	664.89	32.96
camera – gramophon	25.51	706.20	36.12
gramophon – camera	25.51	763.00	33.43
calculator – lamp	26.41	679.42	38.87
lamp – calculator	26.41	843.30	31.32
tape-dispenser – clock	33.82	716.40	47.21
clock – tape-dispenser	33.82	682.70	49.54
case – monitor	34.11	743.15	45.89
monitor – case	34.11	668.00	51.06

Legend: *d* = Distance in meters ; *M* = mean reaction time latencies in ms; *v* = mean velocity in meters per second ($v=d/M$).

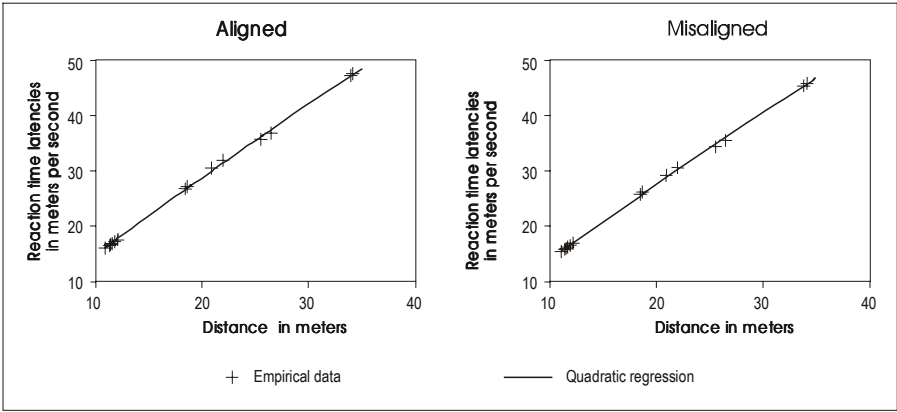


Fig. 3. Regression functions and fit of empirical data

In a third step, the computed regression function was compared to the empirical data. Figure 3 shows the fit of both curves ($F_{aligned}(1,14) = 58113.1$, $P < .0001$; $F_{misaligned}(1,14) = 72262.5$, $P < .0001$).

To evaluate this modelling in a fourth step, I matched the empirical priming effects for each item with the calculated priming effects and computed a correlation coefficient (Spearman’s rho) for the priming effects concerning near and far, aligned and misaligned items. The calculated priming effects are given with table 3. The correlation was $\rho = 0.909$, $P < .001$. This result shows that empirical and estimated priming effects correspond surprisingly well. Approximately 83% of the shared variance are clarified.

5.4 Summary of the Evaluation

To summarize the results, both the fit of the computed regression function and the correlation between the herewith computed priming effects and the empirical priming effects assert the chosen procedure. The introduced flow model approach enables the estimation of empirical priming data when certain assumptions are made. These assumptions were that aligned and misaligned distances are remembered in a different way and that dp/dx is constant but different from zero. This pressure gradient according to the flow direction modulates the resulting velocity function. If this gradient is assigned to zero, the flow changes to a plane Couette flow.

6 Conclusions

The aim of this paper was to provide a new approach to priming processes which can be described as a flow model. A rheological model of the process was illustrated in figure 1. The flow starts with the perception and recognition of one of the objects of a perceived spatial layout (the prime object). The prime object accelerates the

recognition or identification of a second associated object (the target object). The respondents react within a certain period of time. These reaction time latencies have been combined with the relations between prime and target to calculate so called priming velocities. The calculation of those velocities led us to the estimation of aligned and misaligned regression functions.

Table 3. Empirical and estimated priming effects

Item	<i>d</i>	<i>empiric</i>	<i>estimated</i>
lamp – notepad	11.00	-124.85	-92.98
notepad – lamp	11.00	-15.85	-62.39
phone – camera	11.36	-113.65	-92.44
camera – phone	11.36	-96.80	-61.87
camera – monitor	11.52	-124.75	-92.19
monitor – camera	11.52	-47.00	-61.63
tape-dispenser – lamp	11.59	-102.60	-92.08
lamp – tape-dispenser	11.59	-41.95	-61.53
notepad – clock	11.60	-115.65	-92.08
clock – notepad	11.60	-143.15	-61.52
case – phone	11.86	-72.15	-91.68
phone – case	11.86	-124.10	-61.14
gramophone – tape-dispenser	12.07	63.41	-91.37
tape-dispenser – gramophone	12.07	-34.82	-60.84
monitor – calculator	12.10	-64.07	-91.31
calculator – monitor	12.10	0.05	-60.79
camera – notepad	18.42	-147.76	-81.61
notepad – camera	18.42	-80.00	-51.48
phone – lamp	18.64	-52.39	-81.27
lamp – phone	18.64	20.12	-51.16
phone – clock	20.89	-119.10	-77.74
clock – phone	20.89	-30.74	-47.79
case – notepad	21.91	-84.25	-76.12
notepad – case	21.91	-103.76	-46.24
camera – gramophone	25.51	-62.45	-70.39
gramophone – camera	25.51	-5.65	-40.76
calculator – lamp	26.41	-89.23	-68.93
lamp – calculator	26.41	74.65	-39.37
tape-dispenser – clock	33.82	-52.25	-56.75
clock – tape-dispenser	33.82	-85.95	-27.75
case – monitor	34.11	-25.50	-56.27
monitor – case	34.11	-124.85	-27.30

Legend: *d* = Distance in meters. (The priming effect is the difference between mean reaction time and baseline.)

The computation of these regression functions was enabled by constraints which were given by assumption 1 to 8. These assumptions fixed the quantities which were given with the specific solution of the Navier-Stokes equation. In a further step, the regression functions enabled the computation of the priming effects of the modelled

process. These calculated effects could then be compared to the empirical collected priming effects. The resulting correlation was sufficiently high. Therefore, the model seems to be suitable to describe the illustrated priming process which is admittedly restricted to certain conditions.

Yet, three advantages locate the suggested model beyond existing priming theories. First of all, reaction time latencies and priming effects are matched into priming velocities which are consistent to the velocities of the flow field. No further transformation is needed. Secondly, main physical quantities can be related to relevant variables of the mind. One example was the assignment of the fluid to the internal representation. Another quantity is given with a , the distance between the moving and the resting plates. In figure 1 two moving plates are indicated according to the assumption that aligned and misaligned distances are remembered unequally. This distinction, however, is not essential to the suggested flow. Moreover, the distance between plates might also be assigned to other kind of relations and therefore creates the possibility to model also semantic priming processes. Thirdly, the model offers the possibility to understand priming as a non-linear continuous process.

In this sense, the suggested flow model provides the opportunity to describe further priming processes. For this purpose, the quantities given in equation (4) or equation (5) must be varied. The quantity dp/dx , for example, provides a suggestion to demonstrate the efficiency of certain time windows for priming processes. Since it is conceivable that priming effects do not occur when the time window is chosen too short or too long a variation of dp/dx enables the modelling of these discrepancies. To what extent those considerations pass the test of empirical data, is up to further research.

References

- Abraham, R.H. & Shaw, C.D. (1985). *Dynamics – The geometry of behavior* (Part 1: Periodic behavior). Santa Cruz, CA: Aerial Press.
- Anderson, J.R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22, 261-295.
- Anderson, J.R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14, 471-517.
- Beller, H. K. (1971). Priming: effects of advance information on matching. *Journal of Experimental Psychology*, 87, 176-182.
- Birkhoff, G. (1978). *Hydrodynamics*. Westport: Greenwood Press.
- Chown, E., Kaplan, S. & Kortenkamp, D. (1995). Prototypes, location, and associative networks (PLAN): towards a unified theory of cognitive mapping. *Cognitive Science*, 19, 1-51.
- Clayton, K. & Habibi, A. (1991). Contribution of temporal contiguity to the spatial priming effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 263-271.
- Collins, A.M. & Loftus, E.F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Curiel, J.M. & Radvansky, G.A. (1998). Mental organization of maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 202-214.
- De Groot, A.M.B. (1985). Word-context effects in word naming and lexical decision. *The Quarterly Journal of Experimental Psychology*, 37A, 281-297.
- Downs, R.M. & Stea, D.S. (1973) (eds.). *Image and environment. Cognitive mapping and spatial behavior* (pp. 8-26). Chicago: Aldine.

- Franklin, N., Tversky, B., & Coon, V. (1992). Switching points of view in spatial mental models. *Memory & Cognition*, 20, 507-518.
- Gillund, G. & Shiffrin, R.M. (1984). A retrieval model for both cognition and recall. *Psychological Review*, 91, 1-67.
- Hardwick, D.A., Woolridge, S.C. & Rinalducci, E.J. (1983). Selection of landmarks as a correlate of cognitive map organization. *Psychological Reports*, 53, 807-813.
- Janzen, G. (2000). *Organisation räumlichen Wissens. Untersuchungen zur Orts- und Richtungsrepräsentation*. Wiesbaden: DUV.
- Janzen, G., Herrmann, T., Katz, S. & Schweizer, K (2000). Oblique angled intersections and barriers: Navigating through a virtual maze. In C. Freksa, W. Brauer, C. Habel & K.F. Wender (eds.) *Spatial cognition II – Integrating abstract theories, empirical studies, formal methods, and practical applications* (pp. 277-294). Berlin: Springer.
- Kitchin, R.M. (1994). Cognitive maps: what are they and why study them? *Journal of Environmental Psychology*, 14, 1-19.
- Kitchin, R. & Freundschuh, S. (2000). *Cognitive mapping: past, present and future*. London: Routledge Frontiers of Cognitive Science.
- Kuipers, B. (1978). Modelling spatial knowledge. *Cognitive Science*, 2, 129-153.
- Kuipers, B. (1983). The cognitive map: could it have been any other way? In H.L. Pick & L.P. Acredolo (eds.), *Spatial orientation* (pp. 345-359). New York, NY: Plenum Press.
- Levine, M., Jankovic, I.N. & Palij, M. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology: General*, 11, 157-175.
- Luce, R.D. (1986). *Response times. Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.
- Lynch, K. (1960). *The image of the city*. Cambridge, MA: The Technology Press & Harvard University Press.
- May, M., Péruch, P. & Savoyant, A. (1995). Navigating in a virtual environment with map-acquired knowledge: encoding and alignment effects. *Ecological Psychology*, 7, 21-36.
- McClelland, J.L. & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception. Part 1: an account of basic findings. *Psychological Review*, 88, 375-407.
- McKoon, G. & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1155-1171.
- McNamara, T.P. (1986). Mental representations of spatial relations. *Cognitive Psychology*, 18, 87-121.
- McNamara, T.P. (1991). Memory's view of space. In G.H. Bower (ed.), *The psychology of learning and motivation* (pp. 147-186). San Diego: Academic Press.
- McNamara, T.P. (1992a). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1173-1190.
- McNamara, T.P. (1992b). Priming and constraints it places on theories of memory and retrieval. *Psychological Review*, 99, 650-662.
- McNamara, T.P., Halpin, J.A. & Hardy, J.K. (1992). Spatial and temporal contributions to the structure of spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 555-564.
- McNamara, T.P., Hardy, J.K. & Hirtle, S.S.C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 211-227.
- McNamara, T.P. & LeSueur, L.L. (1989). Mental representations of spatial and nonspatial relations. *The Quarterly Journal of Experimental Psychology*, 41 A, 215-233.
- McNamara, T.P., Ratcliff, R. & McKoon, G. (1984). The mental representation of knowledge acquired from maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 723-732.
- Merrill, A.A. & Baird, J.C. (1987). Semantic and spatial factors in environmental memory. *Memory & Cognition*, 15, 101-108.

- Meyer, D.E. & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.
- Milne-Thomson, L.M. (1976). *Theoretical hydrodynamics* (5. ed.). London: The Macmillan Press.
- Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609-626.
- Neely, J.H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner & G.W. Humphreys (eds.), *Basic processes in reading. Visual word recognition* (pp. 264-337). Hillsdale, NJ: Erlbaum.
- Neely, J. H. & Keefe, D. E. (1989). Semantic context effects on visual word processing: a hybrid prospective-retrospective processing theory. In G.H. Bower (ed.), *The psychology of learning and motivation* (Vol. 24, pp. 202-248). New York, NY: Academic Press.
- Pick, H.L., Montello, D.R. & Somerville, S.C. (1988): Landmarks and the coordination and integration of spatial information. *British Journal of Developmental Psychology*, 6, 372-375.
- Posner, M.I. & Mitchell, R.F. (1967). Chronometric analysis of classification. *Psychological Review*, 74, 392-409.
- Presson, C.C., DeLange, N. & Hazelrigg, M.D. (1987). Orientation-specificity in kinaesthetic spatial learning: the role of multiple orientations. *Memory & Cognition*, 15, 225-229.
- Raaijmakers, J.G.W. & Shiffrin, R.M. (1981). Search of associative memory. *Psychological Review*, 88, 93-134.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Ratcliff, R. & McKoon, G. (1981). Does activation really spread? *Psychological Review*, 88, 454-462.
- Ratcliff, R. & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95, 305-408.
- Roskos-Ewoldson, B., McNamara, T.P., Shelton, A.L. & Carr, W. (1998). Mental representations of large and small spatial layouts are orientation dependent. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 215-226.
- Schölkopf, B. & Mallot, H.A. (1995). View-based cognitive mapping and path integration. *Adaptive Behavior*, 3, 311-348.
- Schweizer, K. (1997). *Räumliche oder zeitliche Wissensorganisation? Zur mentalen Repräsentation der Blickpunktsequenz bei räumlichen Anordnungen*. Lengerich: Pabst Science Publishers.
- Schweizer, K. (2001). *Strömt die Welt in unseren Köpfen? Kontiguität und Abruf in mentalen Karten*. (Unpublished habilitation thesis). Mannheim, University of Mannheim.
- Schweizer, K., Herrmann, T., Janzen, G. & Katz, S. (1998). The route direction effect and its constraints. In C. Freska, C. Habel und K.F. Wender (eds.), *Spatial cognition. An interdisciplinary approach to representing and processing spatial knowledge* (pp. 19-38). Berlin: Springer.
- Schweizer, K. & Janzen, G. (1996). Zum Einfluß der Erwerbssituation auf die Raumkognition: Mentale Repräsentation der Blickpunktsequenz bei räumlichen Anordnungen. *Sprache & Kognition*, 15, 217-233.
- Siegel, A.W. & White, S.H. (1975). The development of spatial representations of large-scale environments. In H.R. Reese (ed.), *Advances in child development and behaviour* (pp. 10-55). New York, NY: Academic Press.
- Steck, S. & Mallot, H.A. (2000). The role of global and local landmarks in virtual environment navigation. *Presence*, 9, 69-83.
- Strohecker, C. (2000). Cognitive zoom: from object to path and back again. In C. Freksa, W. Brauer, C. Habel & K.F. Wender (eds.), *Spatial cognition II – Integrating abstract theories, empirical studies, formal methods, and practical applications* (pp. 1-15). Berlin: Springer.
- Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13, 407-433.

- Wagener, M. & Wender, K.F. (1985). Spatial representations and inference processes in memory for text. In G. Rickheit & H. Strohner (eds.), *Inferences in text processing* (pp. 115-136). Amsterdam: North-Holland.
- Werner, S., Krieg-Brückner, B. & Herrmann, T. (2000). Modelling navigational knowledge by route graphs. In C. Freksa, W. Brauer, C. Habel & K.F. Wender (eds.), *Spatial cognition II – Integrating abstract theories, empirical studies, formal methods, and practical applications* (pp. 295-316). Berlin: Springer.
- Werner, S., Krieg-Brückner, B., Mallot, H.A., Schweizer, K. & Freksa, C. (1997). Spatial cognition: the role of landmark, route, and survey knowledge in human and robot navigation. In M. Jarke, K. Pasedach & K. Pohl (eds.) *Informatik'97* (pp. 41-50). Berlin: Springer.
- Zierep, J. (1997). *Grundzüge der Strömungslehre* (6. Aufl.). Berlin: Springer.

Context Effects in Memory for Routes

Karl F. Wender¹, Daniel Haun¹, Björn Rasch¹, and Matthias Blümke²

¹University of Trier, 54286 Trier, Germany

²University of Heidelberg, 69117 Heidelberg, Germany

Abstract. When people experience a new environment they first develop landmark knowledge and second route knowledge. Route knowledge is thought to be different from survey knowledge which may develop with additional experience. The present paper describes three experiments in which participants learned a route through (1) a real maze, (2) a virtual maze, or (3) our university library. Participants were tested for their spatial knowledge using a cued recall procedure. Testing was done either in context, i.e. along the route, or out of context, i.e. in a separate, neutral room.

Results showed a clear context effect. In addition, the context effect generalized along the route in the real environments. However, no generalization was observed in the virtual version. Application of multinomial models revealed that the structure of the knowledge acquired was more complex than assumed by popular models of route knowledge. ¹

1 Introduction

Ever since Tolman [1] coined the term *cognitive map* it has been assumed that people store spatial knowledge in a map-like structure. This concept has also become very popular in disciplines other than psychology. Such a structure, like a cognitive map about a particular environment, does not develop at once, but over time. In an influential paper, Siegel & White [2] proposed three stages of development: (1) landmark knowledge, (2) route knowledge, and (3) survey knowledge. The present paper deals with route knowledge. In particular, we are interested in the base structure of knowledge about routes. We report results from three experiments in which we looked for environmental context effects. The data are also checked for a spatial generalization of context effects. If such a generalized context effect exists, a more complex structure for route knowledge would have to be assumed. A possible candidate for such a theory is the model proposed by Werner, Krieg-Brückner, & Herrmann [3].

In the traditional view route knowledge consists of a mere sequence of landmarks in which each landmark is connected to information of how to get to the next one [2]. This has also been called the *dominant framework* [4]. Route knowledge “... would be ... more akin to paired associate learning, changes in bearing associated with arrival at

¹ **Author Note:** Requests for reprints should be addressed to Karl F. Wender, Department of Psychology, University of Trier, 54286 Trier, Germany. E-Mail: wender@cogpsy.uni-trier.de.

‘stimulus’ landmarks ...”, and “A conservative route learning system would then be, in effect, ‘empty’ between landmarks ...” [2, p. 29]. Many authors have accepted this view. Thorndyke and Hayes-Roth [5] state: “This sequence of prescribed actions may be thought of as a set of stimulus-response pairs or action-condition rules.” Hirtle and Hudson [6] conclude, “Route knowledge is characterized by the knowledge of sequential locations without the knowledge of general relationships”. More recently Gillner and Mallot [7] have described route knowledge as a “structure of association” represented by a *directed graph*, which codes only relations between neighboring landmarks. Similar approaches are also popular with AI-systems. For example, Chown, Kaplan, & Kortenkamp [8] propose a model for route knowledge, called *NAPS*, consisting of nodes corresponding to landmarks where only neighboring landmarks are connected by links. A notable exception is a very elaborate model by Kuipers [9].

The classical model has been questioned since, in particular by Montello [4]. He claims that route knowledge has a more enriched structure from the very beginning. Montello argues that the three forms of knowledge, landmark knowledge, route knowledge, and survey knowledge, develop synchronously right from the beginning. In a strict sense (according to Montello’s view) route knowledge, being a sequence of landmarks and instructions, is an abstraction that possibly occurs in verbal descriptions only, such as in giving directions. Furthermore, we should distinguish between “route knowledge” and “knowledge about routes”. The first is a type of mental representation and the second is a more general knowledge that may include very different aspects. What is at debate here is the format of the knowledge structure involved in following routes.

Although some authors have acknowledged that the classical model is an oversimplification, “... this observation has not resulted, to this point, in any substantial modification to the dominant framework” [4, p. 146]. In this paper we investigate context effects in memory for routes. If such context effects can be found this would be interesting in its own right but at the same time it would show that the dominant framework is overly simplified.

Context effects for route knowledge can be illustrated by the following anecdotal observation: Imagine someone who is familiar with a particular country road, but has not driven down this road for a while. Now, if this person tries to remember how to get from one place to the other via this road, it is not unusual that he or she will not be able to recall the entire route. However, if he or she were to travel the road again suddenly specific details, scenery, objects, directions to take, etc. that could not be recalled while being in a different context might be remembered. We propose that this phenomenon can be explained as a context effect. Objects coming into the range of vision while driving along the road serve as a context facilitating recall. This context effect may even include the activation of details not yet visible from earlier points of observation. We would call this a generalized context effect.

People apparently experience this phenomenon. They report that sometimes they start out on a route without being able to recall all necessary details because they intuitively trust such a context effect. A similar view is expressed by Fukushima, Yamaguchi, and Okada [10] as they introduce a neural network model of spatial memory. Anecdotal evidence also suggests that similar effects may occur in memory for music, where sometimes certain parts of a musical piece can only be remembered if they are preceded by a larger portion of the respective piece.

Context effects in memory experiments have been investigated by many authors (e.g. [11, 12, 13]; for reviews see [14, 15]). A meta-analysis of more than 50 experiments was published recently by Smith and Vela [16]. However, spatial knowledge or route knowledge has not been addressed in any of these papers.

As far as we know, the only other study relevant for context effects in memory for routes is by Cornell, Heth, and Skoczylas [17]. However, the authors did not call it context effects but rather *route expectations*. Accordingly they do not relate their results to the body of research on context effects in the psychological literature. They observed children and adults while walking through a neighborhood. On a second tour, or on the reverse tour, they showed them photographs from scenes along the route. Participants had to judge how confident they were, that the particular scene had occurred on the route. Two results are of particular interest here. The one is, that scenes closer to the participant received a higher rating of confidence than scenes farther away. And the second results is, that scenes ahead on the route were rated higher than scenes in backwards direction. What Cornell et al. [17] call route expectations in our view is a manifestation of a context effect. Furthermore, their comparison of close scenes with farther away scenes resembles what we call a generalized context effect. However, they did not compare an *in context condition* with an *out of context condition*.

By a generalized context effect we mean the following. Along a route you may vary the distance between the immediate context in which something has been learned and the context in which it has to be recalled along a route. In particular, you may select a cue and a target to be recalled along the route in such a way that they are not too far from each other, but separated by some sort of barrier so that they cannot be observed *simultaneously*. If an effect of the context surrounding the cue when recalling the target can still be demonstrated, this is a *generalized context effect* because it generalizes from one context to the next.

In a number of previous studies we tried to find a context effect for route knowledge in a laboratory setting. Participants learned a route on a map presented on a computer screen, which means the stimulus configuration was essentially two-dimensional. In these studies we were able to demonstrate a context effect, but a generalized context effect in the above sense was never found [18].

In the experiments reported below we always contrasted a within context condition with a different context condition. In addition, we used a more controlled set of stimuli that had to be recalled. This, together with our experimental procedure of incremental cued recall, allowed us to more precisely distinguish between immediate and generalized context. And, finally, we applied multinomial models to investigate the internal structure of route knowledge. Such models specify probabilities of recall for particular stimuli given the presence or the memory of other items. By making specific assumptions about these probabilities, different models of the internal structure of route knowledge can be tested. For a recent review on the usage of multinomial models consult Batchelder and Riefer [19].

2 Experiment 1

In this study we tried to find a generalized context effect. To reach this goal, we used an experimental technique called “incremental cued recall”. With this technique

participants first learn a series of stimuli along a route. Then they are given one stimulus as a cue and are asked to recall either the next one, two, or three stimuli. When testing takes place on the route or in a neutral room the difference in recall between these two conditions can be seen as a demonstration of a context effect. Whether an *immediate* or a *generalized effect* is found depends on the particular arrangement of objects and barriers.

2.1 Method

Materials. A life-size maze was built in a lecture room using poster boards. The maze consisted of corridors approximately 1.0 m wide. They were sectioned off completely by opaque plastic above and below the poster boards so that nothing from the surrounding room was visible. The total route was 21 m in length. A floor plan of the maze is shown in Figure 1. There were 6 intersections where participants had to decide between two possible paths. In each instance, one path was a dead end and the other path was the correct continuation of the route.

When approaching an intersection, participants had to stop at a decision point that was marked on the ground by a cross made of white tape. Here the participant had to decide which path to take. Note that participants were not able to look into the dead ends while standing at the decision points.

There were 18 pieces of white legal size paper posted along the walls. Each piece of paper had a word printed on it in large letters. These were the stimuli to be recalled. The stimuli were high frequency words denoting buildings and places, like *bank*, *museum*, *stadium*, *restaurant*, etc. The edges of the sheets were folded in such a way that while standing in front of one sheet, participants could not read the word on the next one. In Figure 1 the decision points are marked by black “plus” (+) signs. The same experimental setup was used in a different study by Mecklenbräuker, Wippich, Wagener, and Saathoff [20].

Procedure and Design. The experiment was divided into a study phase and a test phase. During the *study phase* participants made three trips through the maze together with the experimenter. On the first trip the experimenter informed the subject which route to take when approaching a decision point. During this instruction the words pinned to the walls were mentioned. Subjects were not explicitly instructed to learn these words. Rather, they were told that the words might be of use for finding their way through the maze during later trips. On the second and third trips subjects had to stop at the decision points. They indicated to the experimenter which of the two possible paths they wanted to choose. If necessary they were corrected by the experimenter before they continued through the maze.

In the *test phase*, a cued recall procedure was used. Participants were presented with one of the words from the stimulus set during each trial and were asked to recall either the next one, two, or three stimuli. Testing took place either in the learning environment (i.e., in the maze) or in an adjacent classroom. These are called the *same context* or the *different context* conditions, respectively.

During the test in the same context condition participants traveled through the maze as described above. They were accompanied again by the experimenter who asked them to stop in front of certain, predetermined stimuli. The participants were

then required to recall the next one, two, or three stimuli after which they moved on to the next cue. Stimuli that were to be recalled were never used as cues.

In the different context condition participants were brought to a nearby classroom where they completed the cued recall procedure. Here, the experimenter read the cues to them aloud and showed them pieces of paper identical to those posted in the maze with the cues printed on them. The experimenter wrote down the participant's verbal responses.

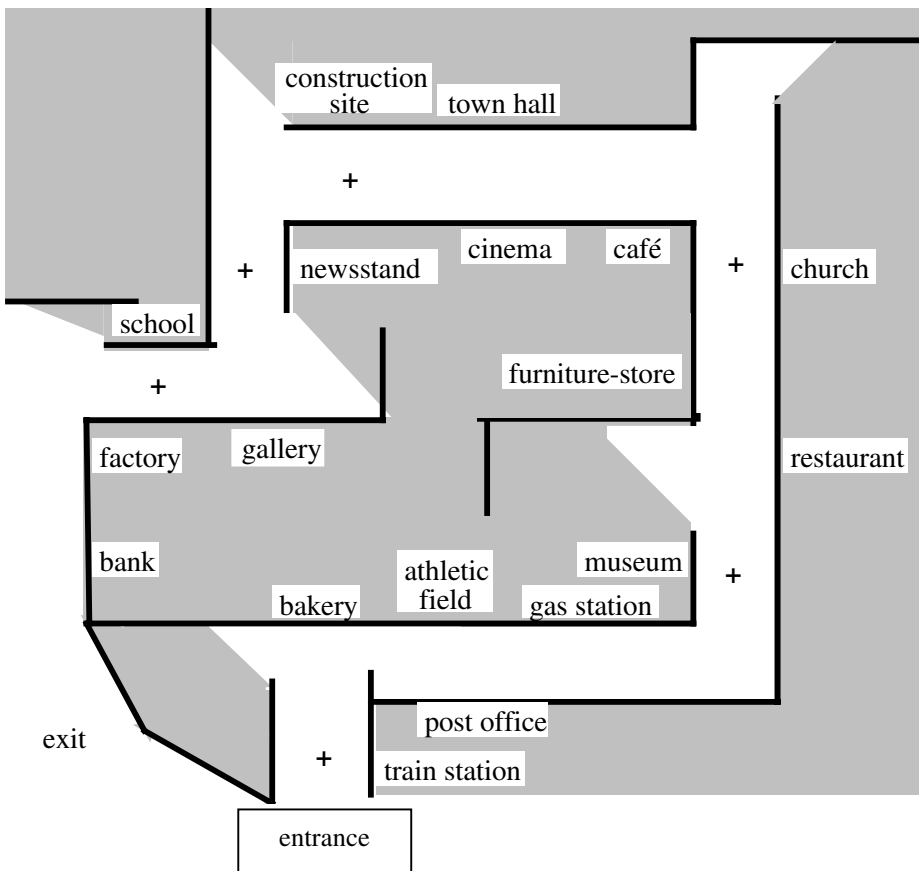


Fig. 1. Floor plan of the maze in Experiment 1. The shaded area represents those parts of the maze that were not visible from inside.

Due to the construction of the maze there were two different stimulus conditions. In the *immediate context condition*, the to-be-recalled stimuli could be seen when standing in front of the cue although they could not be read (because of the folded edges of the paper). Both the cue as well as the to-be-recalled items could be viewed simultaneously, i.e. they were within the same environmental context. In the *generalized context condition*, the last, i.e. the third, of the to-be-recalled stimuli was around the next corner of the maze. Under this condition the cue and the to-be-recalled item could never be seen simultaneously. Therefore, we call these separate

contexts. The question then was, would performance also be better across separate contexts when tested in the maze as opposed to being tested in the classroom. If so, we would call this a generalized context effect.

Participants. Ninety-six students of the University of Trier participated in the experiment. They were paid for their participation.

2.2 Results

Multinomial Analysis. Data were analyzed by a multinomial model and nonparametric measures testing for context effects. We discuss the multinomial models first. For this study, responses were classified into 14 possible response patterns. Figure 2 reports the proportion obtained for each pattern. The response patterns are depicted along the x-axis. For example, as a response to the question “Which stimulus comes next?” one of two responses may occur: correct or false.

In the figure a “1” indicates a correct response whereas a “0” stands for an incorrect answer. If the question was “Which were the next two stimuli?”, four possible response patterns may occur. “1 1” indicates that both stimuli were recalled correctly. “1 0” signifies that the first was recalled correctly, but the second was not, etc. The proportions are displayed for both the same and the different context condition in Figure 2.

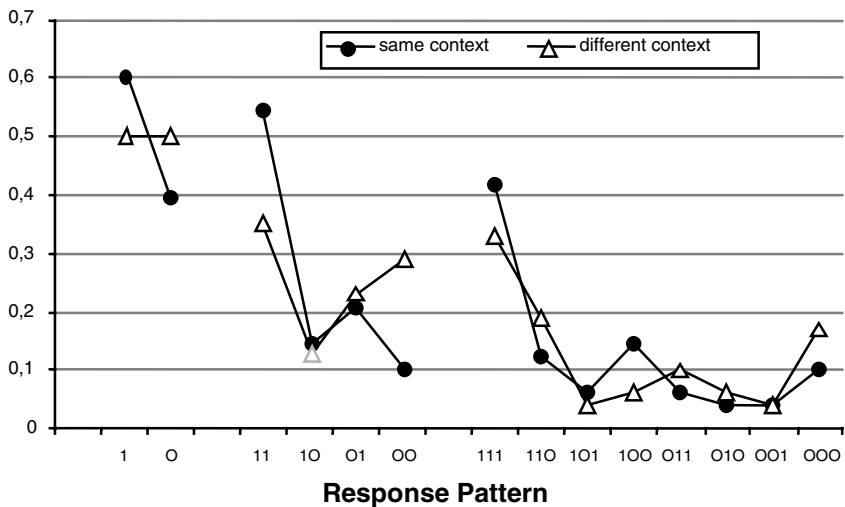


Fig. 2. Experiment 1: Proportions of correct responses.

Several models were applied to the data. Three such models are illustrated in Figure 3. The small circles represent cues and the to-be-recalled stimuli. The links stand for associations between the stimuli. Model M1 states that a link exists from the cue to Stimulus 1 and from Stimulus 1 to Stimulus 2 and so forth. M1 assumes that when a cue is presented, Stimulus 1 can be found with a certain probability. If

Stimulus 1 is recalled, then Stimulus 2 may be found with another certain probability. This model places heavy restrictions on the data. With a probability of $1-p$ the cue does not lead to Stimulus 1. In this case, neither Stimulus 1 nor Stimulus 2 can be recalled. Thus, for example, the expected probability for pattern “0 1” is zero. M1 is called a chain and is our conception of the classical model of route knowledge in the strict sense: There are only links between successive stimuli.

In Model M2 there is a separate link with a certain probability from the cue to each of the stimuli. In other words, there are links from one stimulus to all following stimuli, but only the links from the cue are used. This model has the property that given the cue, each stimulus can be recalled independently. We call this the independence model. It is tested as an alternative to M1.

Finally, Model M3 is the combination of M1 and M2. We call this the full model. According to this model there are alternative routes from the cue to the to-be-recalled stimuli. They can be reached either directly or via one of the other stimuli along the route. Even M3 is not the most general one though: Associations directed backwards are not accounted for. However, such associations are not necessary to describe our data as shown below.

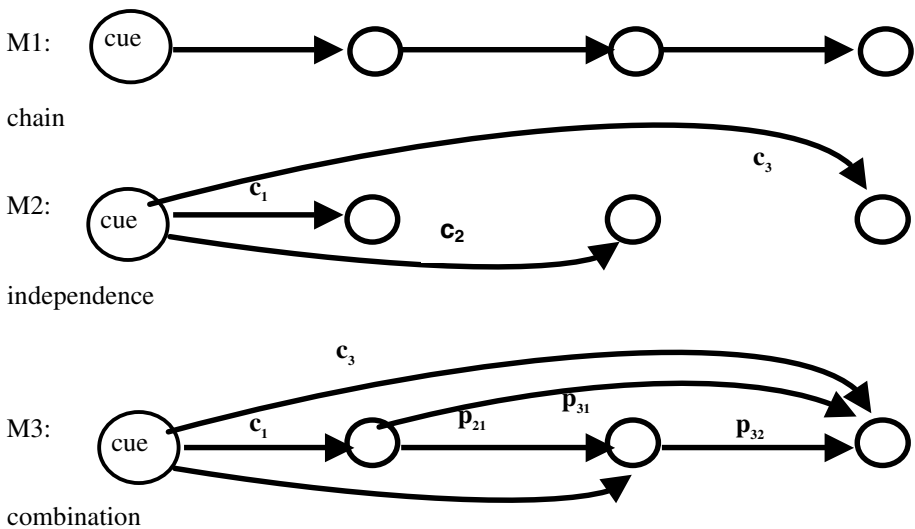


Fig. 3. Graph of the multinomial models in Experiment 1.

From these assumptions, algebraic expressions defining branch probabilities were derived to estimate parameters that would best predict the observed frequencies using the General Processing Tree Algorithm [21]. These expressions are presented in the Appendix for the full model. Equations for the other models were obtained by fixing certain parameters to 1 or 0. Parameter estimation was done by using the routine *AppleTree* developed by Rothkegel [22].

Table 1 includes the statistics for the goodness of fit of the three models and the two recall conditions listed separately. As may be noticed, Models 1 and 2 deviate

drastically from the data, whereas Model 3 fits the data quite closely. For example, the predictions of M3 for the same context condition are shown in Figure 4.

Table 1. Goodness of Fit of the Multinomial Model

	same context			different context		
	χ^2	df	p	χ^2	df	p
M1	312.0	8	<0.0001	348.0	8	<0.0001
M2	17.7	8	0.02	22.4	8	<0.0001
M3	5.1	6	0.54	3.8	6	0.71

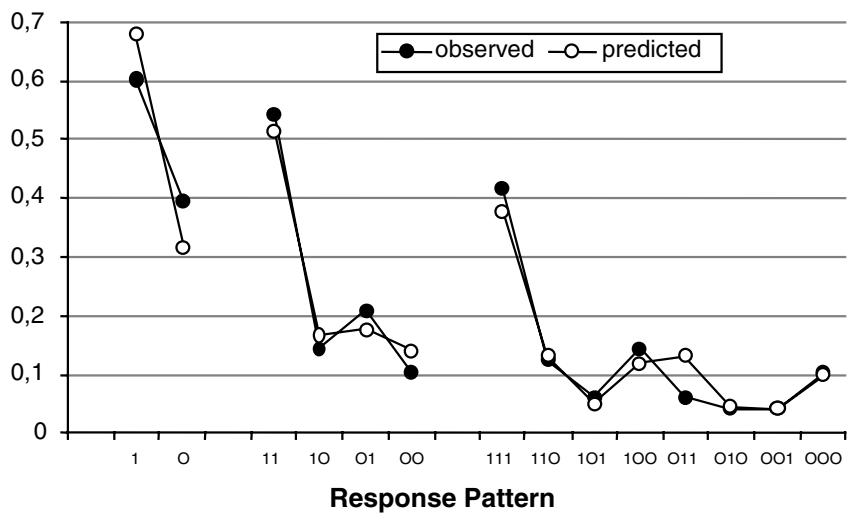


Fig. 4. Experiment 1: Comparison of Model 3 with the data from the same context condition.

Context Effects. Next we tested for context effects. We compared performance conditions under the immediate and the generalized stimulus conditions in the same context and different context. If there is a difference in performance between the same context and the different context condition, this difference can be attributed to a context effect. The results are presented in Figure 5.

Since the data consist of simple frequencies we used a non-parametric Test (χ^2 Test) for analysis. For the immediate context effect, the proportion of correct responses is higher in the same context condition than it is in the different context condition. This difference is reliable ($\chi^2(1) = 6.62, p < 0.025$).

To test whether a generalization of the context effect took place we compared the recall probabilities of the third stimulus “around the corner” under the same context and the different context conditions. When the to-be-recalled stimulus was around the

corner from the cue both stimuli could not be observed from the same position in the maze. When tested in the maze, the proportion of recall for stimuli “around the corner” was higher than when tested in the separate classroom. The difference was reliable ($\chi^2(1) = 4.08, p < 0.05$). This generalized context effect is also shown in Figure 5.

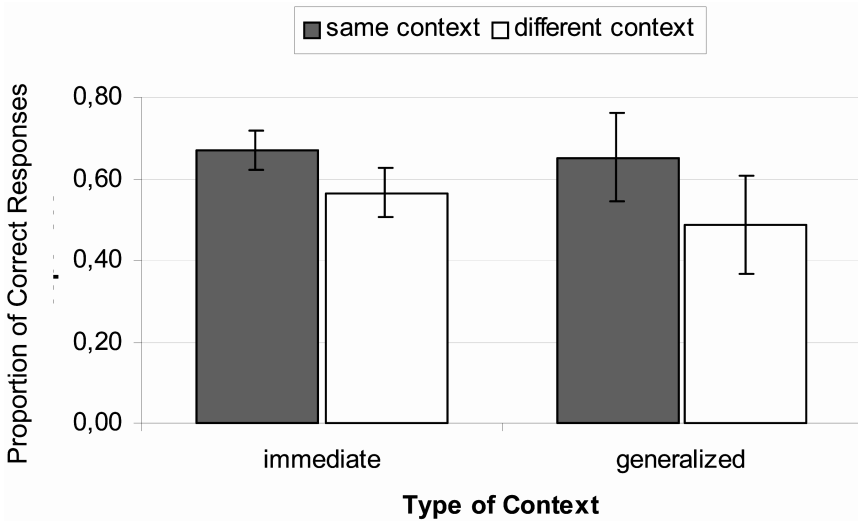


Fig. 5. Context effects in Experiment 1. The error bars show the standard deviations.

2.3 Discussion

Regarding the multinomial analysis, our main result is that M1 does not fit the data. As can be seen in Table 1, M1 deviates highly from the data in both cases and has to be rejected. Yet M1 corresponds with the structure of route knowledge defined as a sequence of landmarks the closest. This sequential structure corresponds with the dominant framework. Hence, we argue that our participants have developed knowledge about the maze that does not resemble route knowledge in its classical sense.

M2 also does not fit the data, at least in the same context condition. The assumption of independent associations from the cue to the other stimuli does not explain the results completely. Finally, M3, being a combination of M1 and M2, describes the results satisfactorily. Therefore, we argue that route knowledge has an internal structure in which successive stimuli are connected by a chain of associations and, in addition, there are associations between stimuli, which do not follow each other successively along the route. M3 has no backward associations, but they are not necessary for the explanation of our data.

We must keep in mind, however, that our participants had made three trips through the maze before they were tested. Three trips are not many, but we cannot exclude the possibility that participants had already begun to develop survey knowledge in the sense of the classical model, possibly developing strict route knowledge on the very first trial. We chose three learning trials because of too low

response probabilities after just one trial. Hence, we can only conclude that already after three trials the internal structure is richer than assumed by the dominant framework for route knowledge. We will return to this point in Experiment 3 below.

Regarding the context effects we have two results. First, route knowledge is susceptible to immediate context effects. Apparently, there were stimuli in addition to the stimulus words that effected enhanced recall. This should be represented in a route knowledge conception. In M3 the effect is incorporated in the parameters for those links that directly connect a cue with the other stimuli.

Second, the context effect includes the stimulus “around the corner”. Stimuli hidden around the next corner could be recalled significantly better in the same context condition as compared to the different context condition. There must have been elements in the current context associated with stimuli included in the next context. Whether this *generalized context effect* is an additive effect or whether it levels off with increasing distance from the cue is a question that cannot be answered by the present data, but requires further research.

3 Experiment 2

In this experiment we tried to replicate the results from Experiment 1 within a virtual reality setting. This is of interest because there is an increasing number of studies using virtual setups to investigate spatial cognition. How virtual environments compare to real ones is a relevant question. There are very few studies in which exactly the same spatial layout has been tested in real and virtual environments. Ruddle, Payne, & Jones [23] used a virtual replication of the environment used in a study by Thorndyke & Hayes-Roth [5]. Their conclusion was, that learning in the virtual environment was comparable to learning in the real situation. Christou & Bühlhoff [24] describe a very elaborate and detailed simulation of real environments but they did not collect data to compare learning in both situations. Richardson, Montello, & Hegarty [25] compared spatial learning in a campus building with learning in a virtual rendering thereof. They found that learning was similar but that a substantial alignment effect occurred in the virtual environment. This suggests that the information used in both situations was different. In Experiment 2 we rebuilt the maze from Experiment 1 in a virtual environment and tested for possible context effects.

3.1 Method

Materials. A virtual version of the maze from Experiment 1 was created on a computer. The virtual version resembled the original one as much as possible. It was built on a Silicon Graphics Machine using the Maya programming environment. The model replicated the floor plan of the real maze. The height of the corridors was determined using the same scale. Photographs of the floor, walls, and ceiling were taken and scanned into the computer for virtual textures. The virtual maze also included the same written stimuli to be recalled.

Procedure and Design. The virtual maze was presented to participants on a 21-inch computer screen as a desktop virtual reality. The maze was shown in QuickTime

Movie form, initiated by the experimenter. The movie stopped at each decision point and the participant had to indicate in which direction to continue. Then the experimenter started the movie again. At each of the to-be-recalled stimuli the virtual camera turned towards the stimulus, stopped for 500 msec, turned into the maze again, and continued along the route. As in the real environment, participants made three trips through the maze during the study phase.

The following test phase was conducted either under a same context or different context condition. In the same context condition participants again watched the movie. The movie stopped at predetermined locations and the camera turned towards one of the stimulus words. This word was used as a cue and participants had to report either the next one, two, or three stimuli. As in Experiment 1, to-be-recalled stimuli were never used as cues and vice versa.

In the different context condition participants were brought to a nearby classroom and the same procedure was used as in Experiment 1.

Participants. Twenty-six University of Trier students participated in the Experiment.

3.2 Results

Proportions of recall were analyzed for the same patterns using the same multinomial models as in Experiment 1. The results are presented in Table 2.

Table 2. Goodness of Fit of the Multinomial Model in Experiment 2

	same context			different context		
	χ^2	df	p	χ^2	df	p
M1	295.0	8	<0.0001	485.0	8	<0.0001
M2	9.7	8	0.29	20.2	8	0.01
M3	5.9	6	0.44	13.3	6	0.04

Context effect results are given in Figure 6. The proportion of correct recall was again higher in the same than in the different context condition for the immediate context effect. However, this difference did not quite reach significance ($\chi^2(1) = 3.57$, $p < 0.10$). For the generalized context effect the data even point in the opposite direction, but are far from being significant ($\chi^2(1) = 0.60$, *ns*).

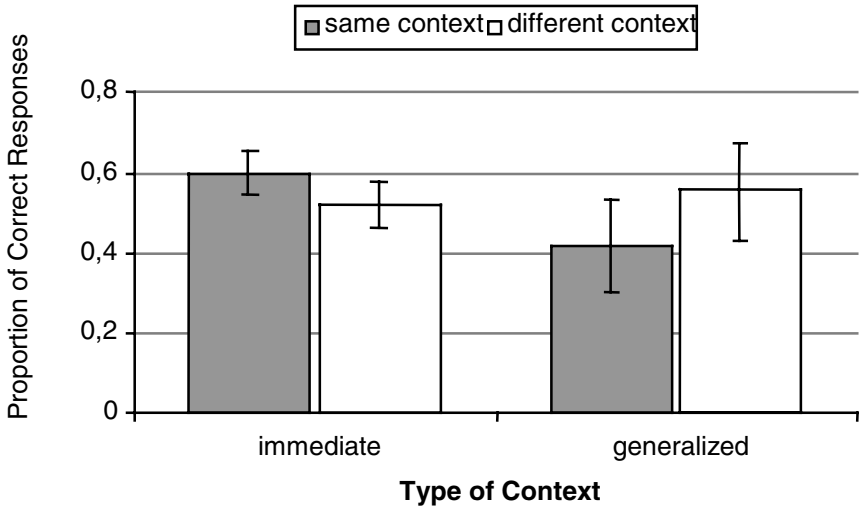


Fig. 6. Context effects in Experiment 2. The error bars represent the standard deviations.

3.3 Discussion

Similar to Experiment 1, M1 deviates significantly from the data. Therefore, a simple chain as a model for route knowledge has to be rejected under virtual environment conditions as well. For M2 and M3, we find a difference in comparison to Experiment 1. In the same context condition M2 does not deviate significantly from the data as it did in Experiment 1. Additionally, even M3 does not fit as it does in all other cases under the different context condition.

In respect to context effects we found a marginal effect for the immediate context. Although this effect is smaller it at least points in the same direction as in Experiment 1. On the other hand, there was nothing close to a generalized context effect. In fact, the results point in the opposite direction, though the difference is far from significant. This is a striking contrast to Experiment 1 where we found a generalized context effect. Hence, we found a difference between real and virtual environments.

In our view, there are three possible reasons for this difference. First, our desktop virtual environment was not sufficiently immersive.

Participants did not feel as though they were in the maze and therefore did not use as many cues as they would have used in a real environment. Another possible explanation is that although we tried to make the virtual reality as realistic as possible we did not quite succeed. The floor, poster boards, and ceiling were much more uniform in the virtual version. A dark gray carpet, for example, covered the floor of the real maze. There were various stains on this carpet which were not represented in the virtual version. Such additional stimuli, perhaps even left unnoticed, may enrich the context in the real situation, and were not available in the virtual environment.

The third reason might be that in the real maze participants moved along the route physically, whereas in Experiment 2 participants were sitting on a chair. It is possible that this lack of proprioceptive feedback may have contributed to the results.

4 Experiment 3

This experiment was conducted to apply our results to a more realistic environment. We used a route in an even more realistic setting, at least for members of a university. Also, we used additional dependent measures to access spatial knowledge. In the cued recall procedure we not only asked for stimuli to be remembered, but we also asked which turn should be taken at the next decision points. In addition, we included some questions to control whether survey knowledge had been developed. Data were again analyzed using multinomial models, and we looked for possible context effects.

4.1 Method

Materials. A route was constructed using the corridors defined by the bookshelves and other local objects in the university library. Trier's university library has a rather complex spatial layout extending over four large buildings connected by covered walkways and including several hundred yards of paths. Our route included 39 decision points, where participants had to decide which of two alternative turns to take. Decision points were indicated by a stripe of brown tape on the floor. Figure 7 depicts a floor plan of the route. The total route length was about 450 m.

A legal size sheet of white paper was posted at each decision point with a word printed on it in large letters accompanied by a pictogram. These were the stimuli to be remembered. The pictograms were added to improve memory performance. All stimuli were high frequency words denoting buildings and places located in a typical town (like bank, bridge, bakery, pharmacy etc.). Each stimulus was clearly separated from the background as they were each framed by a large pink oval. The route was constructed in such a way that the participants were not able to see any of the following stimuli while standing at a particular decision point. Thus, in this experiment the second and third to-be-recalled stimuli always belonged to the next context.

Participants. Sixty freshmen of the University of Trier participated in the experiment. It was made sure in advance that they had no prior knowledge of the spatial layout of the library. They were paid for their participation.

Procedure and Design. Prior to the study phase participants were brought into the main lobby of the library to receive verbal instructions and complete a pretest. We applied the subscale for spatial abilities from the Wilde Intelligence Test [26]. However, the results showed no systematic relationship to the main results of Experiment 3. Therefore we do not report the results here.

In the study phase participants were led along the route twice in groups of two. An experimenter accompanied each group. Participants were instructed to learn the stimuli because this might help them to find their way along the route on future trips.

During the first trip of the study phase the group was stopped at every decision point. Participants were instructed to associate the stimulus with the correct direction to be taken. One trip along the route lasted approximately 12 minutes. During the second trip participants stopped at each decision point and indicated to the experimenter which of the two possible paths they wanted to choose.

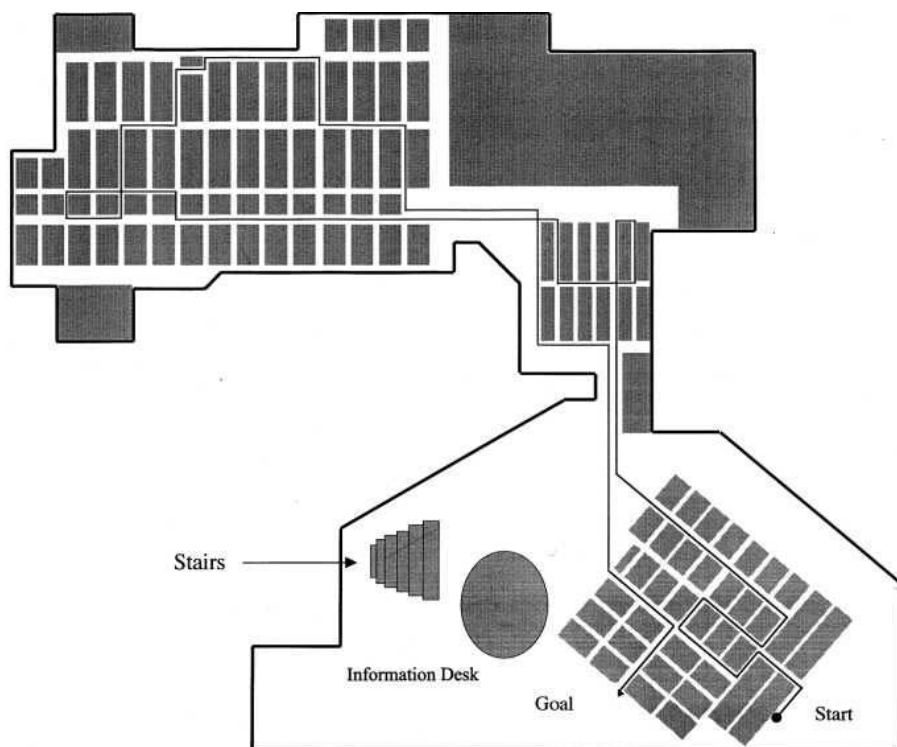


Fig. 7. Floor plan used in Experiment 3. The gray fields represent book shelves and served as visual barriers.

Participants were tested by the same cued recall procedure as used in the previous experiments. In the *same context condition* participants walked along the route again and had to respond at certain decision points. In the *different context condition* participants were brought to a nearby classroom outside of the library where they answered the same questions.

Participants were tested individually. One participant of each pair was randomly assigned to the same context condition and the other one to the different context condition. All participants were brought outside the library for the same amount of time to keep time conditions constant and to keep all context changes equal other than those intended.

After the instructions were given in the lobby and the study phase had been completed, participants in the same context condition were guided to predetermined locations, i.e. decision points on the route. These points and the stimuli posted there served as recall cues. Standing in front of a particular stimulus, participants were required to recall either the next one, two, or three stimuli. The question types were randomly distributed along the route.

Secondly, while standing at a certain decision point participants were asked to recall which turn to take at the next, next two, or next three decision points.

So, at each decision point participants had to answer two questions, one about the to-be-recalled stimulus words and one about the possible turns. After that, they moved on to the next cue. Stimuli to be recalled were never used as cues.

In addition, participants had to give a third response at four specially selected decision points. This was done to test for survey knowledge. Participants were asked to *estimate* the *bearing* from where they were standing to the location of four distant, nonvisible stimuli. These stimuli were selected from very different regions of the route. We argue that the ability to point correctly towards a distant, nonvisible location proves that this person has at least some survey knowledge. Participants were given a sheet of paper with a circle printed on it. They were instructed that the center of the circle is the location where they were standing. The top of the paper was to be aligned with the direction they were facing. Participants were allowed to look around. Finally, the estimated bearing towards the stimuli specified by the experimenter was to be indicated by a little mark on the circle's perimeter.

In the different context condition participants completed the cued recall procedure in a nearby classroom. During each trial the experimenter read one of the cues aloud and presented a copy of the cue as it had been posted on the route. As in the same context condition, participants had to recall either the next, next two, or next three stimuli. Also, they had to recall which turn to take at the next, next two, or next three decision points. Bearing estimates were also obtained at four decision points. After answering the questions, participants were shown all the stimuli that would have to have been included in the correct answer. This was done, because in the same context condition participants could see the correct stimuli when walking towards the next cue. To keep things equal, these stimuli were also shown in the different context condition.

4.2 Results

Resulting from the cued recall procedure in this experiment, we have two dependent measures: recall of stimulus words and recall of turns. Figure 8 shows the results for both measures.

The proportions of correct recall of the next, second, or third stimulus word and direction statement of both context conditions are displayed. Correct responses for the second or third turn are included independent from answers to previous stimuli. First, it is apparent that the proportion of correct responses is higher for turns than for words. This, however, is an artifact of the procedure because the number of possible alternatives is much smaller for turns than for words.

Data were also analyzed for context effects. The results show that under all conditions proportions of correct responses are higher in the same context than in the different context. This was the case for words ($\chi^2(1) = 31.5, p < 0.001$) as well as for turns ($\chi^2(1) = 90.4, p < 0.001$). Analyzed separately, the context effect was significant for words for the first ($\chi^2(1) = 19.2, p < 0.001$) and the second item ($\chi^2(1) = 14.5, p < 0.001$), but the context effect was not significant ($\chi^2(1) = 0.74, p < 0.39$) for the third item of this dependant measure. For turns, on the other hand, the effect was significant for all three types of items. The χ^2 values were 42.2, 41.6, and 12.0 respectively. Thus, we have an immediate context effect for words and for turns,

which generalizes along the route. The effect is stronger for turns, including the third stimulus. The third stimulus is not reached by the generalization for words.

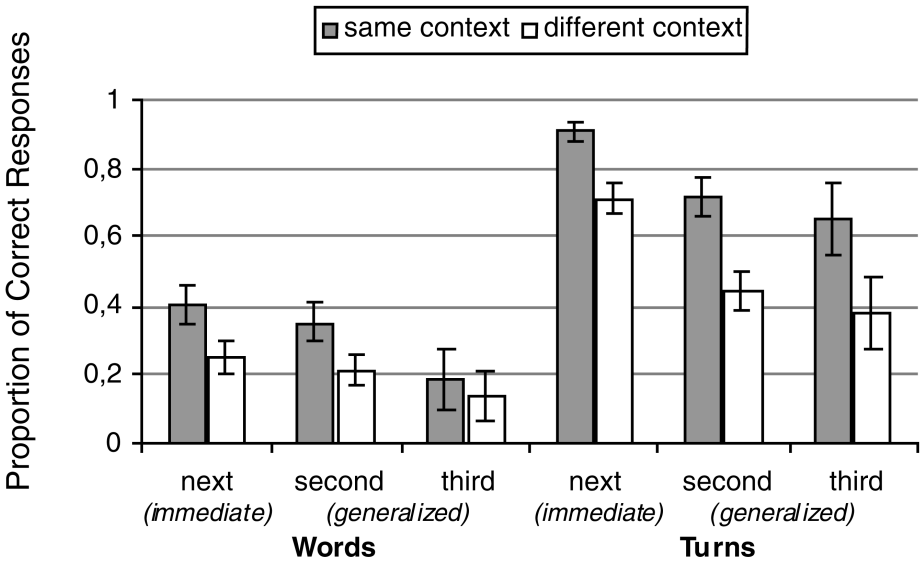


Fig. 8. Proportion of correct recall of words and turns in Experiment 3. The error bars give the standard deviations.

Next, the multinomial models were again fit to the data, the results of which are given in Table 3. As in the previous experiments neither M1 nor M2 fits the data. M3 produces an adequate prediction for turns. For words, however, even M3 does not fit under the same context condition. Therefore, an additional model, M 4, was applied to the data. In this model for different types of questions (i.e. asking for the next, next two, or next three stimuli), parameters between the stimuli were allowed to take on different numerical values. This model fits the data best.

As discussed above, the context effect spreads from one local context to the next one. This is reflected in the parameter of the model that fits the data best. For example, the parameter c_3 in M3 (see Figure 3) may be interpreted as representing the effectiveness of the cue in a particular context. For turns, this parameter was estimated to have the value 0.50 in the same and 0.27 in the different context condition. For words, where we did not find a generalized context effect, the respective values were close to zero (0.02 and 0.09).

Finally, we analyzed the bearing estimations to distant stimuli. Because these are data on a circular scale they were analyzed using circular statistics [cf. 27]. We first computed the *mean vector* of all individual estimations. The direction of the mean vector gives the mean estimated bearing. The length of the vector reflects the variability of the distribution. If the individual estimates are distributed evenly in all directions, the length of the mean vector is close to zero. If all individual estimates point in the same direction, the mean vector is of maximum length.

Table 3. Experiment 3: Goodness of Fit of the Multinomial Model

Turns						
same context				different context		
	χ^2	df	p	χ^2	df	p
M1	291.2	8	<0.0001	594.3	8	<0.0001
M2	38.0	8	<0.0001	25.0	8	0.0016
M3	3.9	5	0.56	6.1	5	0.30
M4	0.27	1	0.61	1.2	1	0.26

Words						
same context				different context		
	χ^2	df	p	χ^2	df	p
M1	373.4	8	<0.0001	331.9	8	<0.0001
M2	74.1	8	<0.0001	20.1	8	0.0098
M3	19.9	5	0.0013	4.8	5	0.44
M4	0.88	1	0.35	0.02	1	0.90

The situation is illustrated in the following diagram. Figure 9 shows two circular histograms, one for the same context condition and one for the different context condition.

The four directions that had to be judged under each condition are represented together in each of the two diagrams. The data are combined in such a way that the correct directions always point “north”, i. e. zero.

The size of the shaded areas corresponds with the number of participants pointing in a particular direction. As may be noticed there is quite some variation between participants. Some people even pointed in the opposite direction.

The black arrows in Figure 9 represent the mean vectors. The length of the mean vector was 0.48 under the same context condition and 0.34 under the different context condition. The length of the mean vector can be tested against zero using the Rayleigh test [27, p. 54], a test against uniform random distribution. The test statistic of the Rayleigh Tests z assumed the following values: 5.76 ($p < 0.01$) in the same context and 3.12 ($p < 0.05$) in the different context. This means that there is a substantial covariance between estimates. This does not mean, however, that the estimations were in the correct direction.

If we want to consider both the variation between judgements as well as the correctness of the mean direction we can compute what has been called the *homeward component* in animal research [27, p. 15]. This measure is given by the projection of the mean vector on to the *correct vector*, normalized to unit length. The homeward component is called v . In Figure 9 the homeward component is represented by the thick vertical lines. The measure v combines both the variation in judgements, via the length of the mean vector, as well as the deviation from the correct direction, via the angle between the mean vector and the correct direction. The numerical estimates for

v were 0.45 for the same context condition and 0.29 for the different context condition. V can be tested against zero using the u -statistic provided by Batschelet [27, p. 59]. This statistic was significantly different from zero in both cases: $u = 3.17$ ($p < 0.001$) in the same context and $u = 2.14$ ($p < 0.05$) in the different context condition.

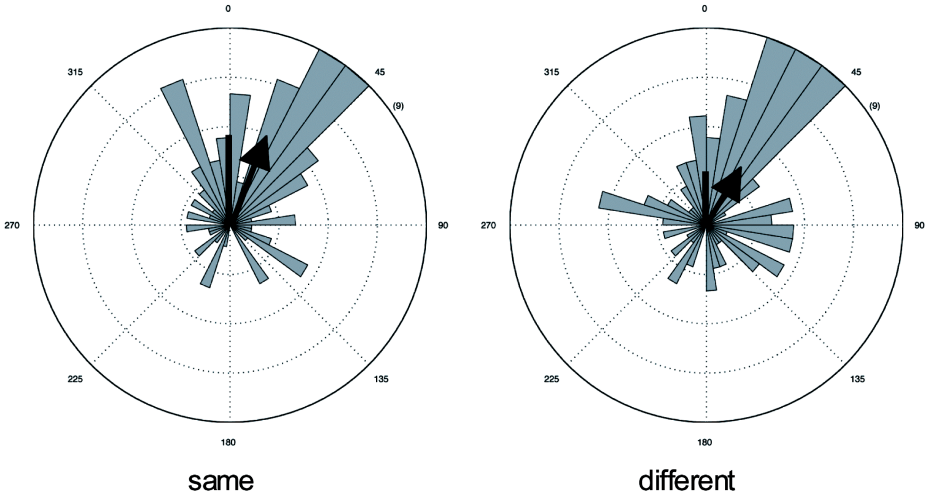


Fig 9. Circular histograms of bearing estimates for the same and the different context condition. The thick black arrows represent the mean vectors. The vertical thick lines are the homeward component.

A comparison of the two diagrams in Figure 9 reveals that the variation is larger in the different context condition than in the same context. This is indicated by the length of the mean vector as well as by the shaded segments of the circular histogram. Furthermore, the direction of the mean vector is closer to the correct direction “north” in the same context condition. This is reflected by the numerical values obtained for v under both conditions. The homeward component can, theoretically, vary between zero and one. The maximum value is reached if all people point in the same correct direction unanimously. The component achieves zero if all people point in different directions resulting in the mean vector’s length being equal to zero, or if the mean vector is 90 degrees or more off the correct direction. Varying between zero and one, the homeward component may be interpreted as measuring the correctness of the pointing behavior.

As an additional analysis following a proposal by Montello, Richardson, Hegarty, & Provenza [28], we computed a *constant error* and a *variable error*. The constant error is the signed difference between the mean direction of estimations and the correct direction. It is a measure of how correct the estimates are on average. According to our data, the constant errors were 22.2° in the same context condition and 30.05° in the different context condition. This means that there was a 7.85 degree difference between the same context condition and the different context condition. Finally, we computed the average of the unsigned differences between the estimations and the mean direction. This is called the variable error by Montello et al.. This

variable error was smaller in the same context condition (46.3°) than in the different context condition (60.42°). This difference is reliable ($F_{(1,45)} = 4,389$, $p < 0.05$). The effect is small ($f = 0.02$), but it is another manifestation of a context effect.

4.3 Discussion

In line with the results from Experiments 1 and 2, the simple models M1 and M2 did not fit the data. This is true for both dependent measures. The same results occurred in the same context as well as in the different context condition. Thus, we have again strong evidence against a conception of route knowledge in the classical sense.

In most cases the full model (M3) actually fits the best. However, for the recall of words in the same context condition, M3 also deviates significantly from the data ($\chi^2(5) = 19.9$, $p < 0.0013$). Though M3 is less restrictive than M1 and M2, it still has restrictions. For example, all possible links from a cue towards the to-be-recalled items have parameters of equal numerical value, regardless whether participants were asked for one, two, or three stimuli. It is conceivable, on the other hand, that a cue may have different power, depending on the recall task. For example, when asked for just one single item participants may have been more confident about completing this task and therefore be more successful as compared to when they were asked about the first item of a sequence of three. This is in fact found in the data. The biggest difference between the empirical data and the probabilities predicted by M3 is shown in the correct recall of one single stimulus. Here, participants were performing better as would be expected from the overall fit of the model. Hence, a less restrictive model (M4) was generated by using different parameters for the three different tasks. We expected not only a better fit of M4, but also a decline in numerical value of the parameters for the next, next two, or next three stimuli. This was the case.

Similar to Experiments 1 and 2 we found a context effect. And again this context effect spreads along the route. However, this generalization is stronger for turns than for words. We derive two conclusions from our results. First, turns probably were more important for our participants than words because turns are more essential for following the route. Second, the context effect for words seems to level off with increasing distance. That is, the generalization gradient has a negative slope. This was not necessarily expected. An alternative could have been that the context effect constitutes an additive constant adding to all recall probabilities when in the same surrounding.

Finally, there is the question of how much survey knowledge our participants had developed. On the one hand, the direction judgements are not completely at random as indicated by the length of the mean vectors. Also, the homeward components are different from zero. On the other hand, the values obtained are far below the maximum value possible. We conclude from our data that a substantial amount of survey knowledge developed, but that this knowledge is far from being complete. The occurrence of this after only two trips along a rather complex route is in agreement with Montello's claim that route knowledge rarely exists in the pure form, but survey knowledge develops right from the beginning. In addition, the values show that the amount of survey knowledge is higher under the same context condition. This is another manifestation of a context effect.

5 General Discussion

In the studies presented here participants were required to learn their way through a maze or a route through a highly complex building. Afterwards, they were tested by cued recall either on the route or in a separate, neutral room. We found higher recall rates when people were tested in the same context. Such a context effect may be interpreted as the existence of associations between the to-be-recalled items and elements of the surrounding situation. We assume that the stimuli that had to be learned and were used as cues serve as landmarks. The context effects, in particular the generalized context effect, suggest that the representation between landmarks was not “empty”, as proposed by Siegel & White [2]. The context effect implies that there were elements and aspects in the situation and associations between them that enhance recall in the same context condition. These associations cannot be recalled intentionally when the person is in a different room, apparently. Since these elements, although possibly not consciously acknowledged, are effective they should be represented in a model of route knowledge. Route knowledge would then consist not only of associations between landmarks (two by two), but also of higher order associations mediated by context elements.

In conclusion, our results from three experiments, with only two or three trips along the route, speak against the conception of route knowledge as a mere sequence of landmarks. A more multifaceted structure has to be assumed for the representation in memory. The internal structure is more complex than that of a chain and, furthermore, context effects document additional associations between landmarks and their surroundings. This was found in a laboratory set-up, but also in a realistic environment. Thus, the theory about mental maps or memory for spatial relations has to take this into account. In a recently published paper Werner et al. propose the notion of a route graph for modeling navigational knowledge. In their view, “A route is a concatenation of directed Route Segments from one Place to another” [3, p. 305]. The conception of a “*place*” here is broader than that of a landmark: “... the notion of a Place has different implications depending on the scenario at hand”. Thus, we assume that in principle, context elements can be incorporated into the representation of a place. But the details still require further investigation.

As a side effect, we found that these context effects were stronger in the real maze than in the virtual version. This can perhaps be attributed to the fact that our desktop virtual reality, although of high fidelity, was significantly less immersive. And, perhaps even more important, proprioceptive stimuli are absent in a desktop presentation. This question requires further research.

Acknowledgments

The study was supported by the Deutsche Forschungsgemeinschaft under grant We/498 27-2. We would like to thank Monika Wagener for developing the spatial layout used in Experiments 1 and 2. Furthermore, we would like to thank Claus C. Carbon, Albert-Georg Lang, and Sabine Schumacher for their assistance in data collection and Rainer Rothkegel for his assistance in data analysis.

Also, we would like to thank Prof. F. N. Rudolph from the Fachhochschule Trier for his support in programming the virtual environment. And finally we would like to thank Erin Marie Thompson for checking our English.

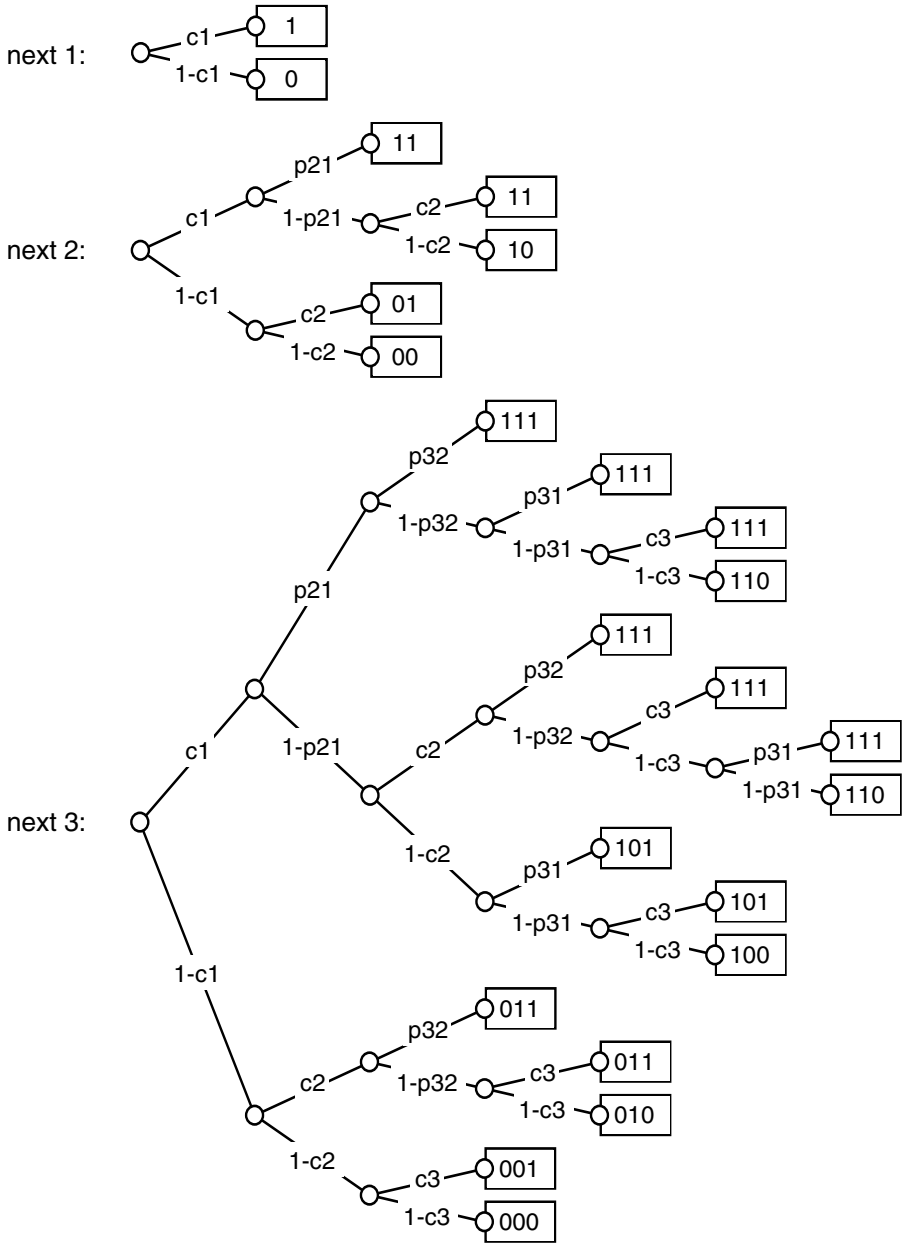
References

1. Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55, 189-208.
2. Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In H. W. Reese (Ed.), *Advances in Child Development and Behavior* (pp. 9-55). New York: Academic Press.
3. Werner, S., Krieg-Brückner, B., & Herrmann, T. (2000). Modeling navigational knowledge by route graphs. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Spatial cognition II* (pp. 295-316). Berlin: Springer.
4. Montello, D. R. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments. In H. Egenhofer, & R. Golledge (Eds.), *Spatial and temporal reasoning in geographic information systems* (pp. 143-154). Oxford: Oxford University Press.
5. Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14, 560-589.
6. Hirtle, S. C., & Hudson, J. (1991). Acquisition of spatial knowledge for routes. *Journal of Environmental Psychology*, 11, 335-345.
7. Gillner, S., & Mallot, H. A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience*, 10, 445-463.
8. Chown, E., Kaplan, S., & Kortenkamp, D. (1995). Prototypes, location, and associative networks (PLAN): Towards a unified theory of cognitive mapping. *Cognitive Science*, 19, 1-51.
9. Kuipers, B. (1978). Modeling spatial knowledge. *Cognitive Science*, 2, 129-153.
10. Fukushima, K., Yamaguchi, Y., & Okada, M. (1997). Neural network model of spatial memory: Association recall of maps. *Neural Networks*, 10, 971-979.
11. Bower, G. H. (1981). Mood and memory. *American Psychologist*, 36, 129-148.
12. Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66, 325-332.
13. Tulving, E., & Thomson, D. H. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 359-380.
14. Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies, & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13-34). New York: Wiley.
15. Smith, S. M. (1994). Theoretical principles of context-dependent memory. In P. Morris & M. Gruneberg (Eds.), *Theoretical aspects of memory* (Aspects of Memory, 2nd ed., Vol. 2, pp. 168-195). New York: Routledge.
16. Smith, S. M. & Vela, E. (2001). Environmental context-dependent memory: A Review and meta-analysis. *Psychonomic Bulletin & Review*, 8, 203-220.
17. Cornell, E. H., Herth, C. D., & Skoczylas, M. J. (1999). The nature and use of route expectancies following incidental learning. *Journal of Environmental Psychology*, 19, 209-229.

18. Schumacher, S., Wender, K.F., & Rothkegel, R. (2000). Influences of context on memory for routes. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Spatial cognition II* (pp. 348-362), Berlin: Springer.
19. Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6, 57-86.
20. Mecklenbräuker, S., Wippich, W., Wagener, M., & Saathoff, J. E. (1998). Spatial information and actions. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial Cognition* (pp. 39-61). Berlin: Springer.
21. Hu, S., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, 59, 21-47.
22. Rothkegel, R. (1999). AppleTree : A multinomial processing tree modeling program for Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 31, 696-700.
23. Ruddle, R.A., Payne, S.J., & Jones, D.M. (1997). Navigating buildings in "desktop" virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied*, 3, 143-159.
24. Christou, C., & Bühlhoff H. H. (2000). Using realistic virtual environments in the study of spatial encoding. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Spatial cognition II* (pp. 317-332). Berlin: Springer.
25. Richardson, A. E. Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition*, 27, 741-750.
26. Jaeger, A. O., Althoff, K. (1983). *Der Wilde-Intelligenz-Test (WIT)*. Goettingen: Hogrefe.
27. Batschelet, E. (1981). *Circular statistics in biology*. New York: Academic Press.
28. Montello, D.R., Richardson, A.E., Hegarty, M., & Provenza, M. (1999). A comparison of methods for estimating directions in egocentric space. *Perception*, 28, 981 – 1000.

Appendix

Decision trees for the multinomial models M1, M2 and M3. The top tree contains the parameters that determine the probabilities for the possible responses to the question: "Which was the next stimulus?" The middle tree gives the parameters for the response patterns to the question: "Which were the next two stimuli?" And the bottom tree contains the parameters for the question: "Which were the next three stimuli?" To obtain the probabilities the parameters along the branches have to be multiplied and the products have to be added for identical response patterns. For the models M1 and M2 certain parameters have to be fixed to 0 or 1.



Towards an Architecture for Cognitive Vision Using Qualitative Spatio-temporal Representations and Abduction

Anthony G. Cohn, Derek R. Magee, Aphrodite Galata,
David C. Hogg, and Shyamanta M. Hazarika

School of Computing, University of Leeds, Leeds, LS2 9JT, UK
{agc,drm,afro,dch,smh}@comp.leeds.ac.uk

Abstract. In recent years there has been increasing interest in constructing cognitive vision systems capable of interpreting the high level semantics of dynamic scenes. Purely quantitative approaches to the task of constructing such systems have met with some success. However, qualitative analysis of dynamic scenes has the advantage of allowing easier generalisation of classes of different behaviours and guarding against the propagation of errors caused by uncertainty and noise in the quantitative data. Our aim is to integrate quantitative and qualitative modes of representation and reasoning for the analysis of dynamic scenes. In particular, in this paper we outline an approach for constructing cognitive vision systems using qualitative spatial-temporal representations including prototypical spatial relations and spatio-temporal event descriptors automatically inferred from input data. The overall architecture relies on abduction: the system searches for explanations, phrased in terms of the learned spatio-temporal event descriptors, to account for the video data.

1 Introduction

There has been extensive research into techniques for Computer Vision (CV), but much of this has concentrated on important, but low level methods. Although these low level techniques can sometimes be applied directly in a system, in general, a more high level understanding of the scene will be required. The relative paucity of research in this area¹ has resulted in a number of EU funded projects on *Cognitive Vision* which allow a much greater “semantic” access to and processing of visual information. The University of Leeds is a partner in one such project, CogVis (Cognitive Vision Systems, IST-2000-29375). This paper describes our approach to the goal of creating a cognitive vision system, and in particular the combination of qualitative spatial reasoning techniques with more conventional CV research.

First, it is worthwhile quoting from the Technical Annexe of CogVis to give a “definition” of cognitive vision:

¹ Though see, e.g. [1,2,3,4].

“Considering the general definition of cognition as “the process of knowing, understanding and learning things” it is possible to derive some key characteristics for cognitive vision:

- Vision is a process that operates in a spatio-temporal context. I.e. vision is not instantaneous, it evolves over time and incorporates information to generate “answers”.
- Vision uses and generates knowledge (that includes information that is not organized spatially). This implies that a fundamental part of studies of visual processes is consideration of representations and memory.
- The visual process generates/maintains models of the environment in terms of its geometry, and semantic labels for events and entities in the environment. I.e. “understanding” implies an ability to generate an explicit description of the perceived world in terms of objects, structures, events, their relations and dynamics that can be used for action generation or communication.
- Learning implies an ability to generate open-ended models and representations of the world. That is, the model of the system and its use cannot be based on a closed world assumption, but rather on a model that allows automatic generation of new representations and models.
- Vision is a process which implies that it operates in the context of an “agent” that provides a task context and has finite resources in terms of computation, memory and bandwidth.”

Thus key to our approach will be the integration of learned models, of explicit spatio-temporal representations and open ended reasoning allowing the generation and explicit manipulation of symbolic hypotheses.

2 The Traditional Computer Vision Approach

Traditionally, CV approaches the problem of scene understanding as one of finding methods to transform between input images or sequences and an N dimensional parameterisation (where N is arbitrary but fixed) [5,6,7,8,9]. Ad-hoc methods are often used for representing and understanding multiple objects. Many spatial and temporal classification and prediction methods (e.g. Hidden Markov Models) have been developed based on this ‘world in N dimensions’ paradigm [10,11]. The problem is that the world is not generally well described by an N dimensional parameterisation (although the higher the N , the better the description generally), rather as the sum of a number of concepts. Methods attempt to reduce the computational dimensionality without reducing the parameterisation dimensionality (for reasons of computational efficiency) by using dimensionality compression techniques such as Principal Components Analysis or Vector Quantisation. This does not however solve the problem that a fixed N dimensional representation is not a good way of representing a general scene. CV methods generally *approximate* the real world and thus are rarely 100 % *accurate*.

3 Qualitative Spatio-temporal Representations and Reasoning

The development of Qualitative Spatial Reasoning (QSR) [12] has been driven by the realisation that much cognitive representation and processing of spatial data is qualitative – e.g. most everyday natural language spatial expressions are purely qualitative (“on the table”, “behind the tree”, “in the bottle”) and moreover that much uncertainty in spatial data can be abstracted away through the use of qualitative representations that discretize a continuous space into a finite and small number of *relevant* possibilities – qualitative representations are typically abstract but accurate (rather than precise and possibly inaccurate). Thus, for example, the RCC-8 calculus [13] has eight jointly exhaustive and pairwise disjoint relations categorising possible topological relations between a pair of regions – see figure 1; a very similar calculus has also been derived from alternative semantic primitives [14]. Indeed, the use of regions as a primitive spatial entity (rather than points) also helps abstract away from uncertainty. If the boundary of real world regions is unknown or in some other way indeterminate, then an extension of the calculus has been designed to handle such regions [15] (see also [16]). Other QSR calculi have been designed to represent and reason about orientation (e.g. [17,18,19]), convexity [13], shape (e.g. [20]) and congruence [21,22]. An important notion when considering dynamic spatial knowledge is that of a *continuity network* or *conceptual neighbourhood* which specifies which relations are neighbours as objects move or transform continuously over time as this allows for prediction and explanation of spatio-temporal data – see figure 1. For a survey of QSR see [12] or [23].

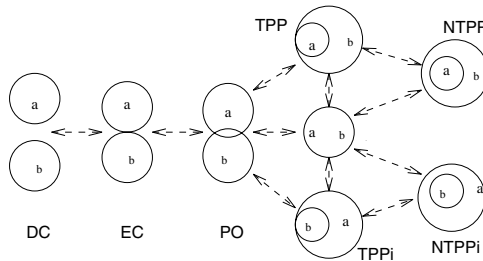


Fig. 1. 2D illustrations of the relations of RCC-8 calculus and their continuous transitions (*conceptual neighbourhood*).

When reasoning with qualitative spatial data over time, one possibility is to take a ‘snapshot’ viewpoint, and describe dynamic behaviour as a set of temporal states, where each state consists of a qualitative spatial representation and their temporal relationship described by a temporal logic. This approach has been extensively investigated by [24,25,26] and a number of useful complexity results are given. An alternative approach is to view the world as spatio-temporal

histories [27] and extend purely spatial qualitative representation languages to qualitative spatio-temporal languages with relations which hold between such space-time histories [28,29,30].

To apply QSR methods to CV requires a qualitative description of the world/scene as an input. In constructing this description, the system abstracts away from the initial (potentially erroneous) data and thus may remove error components in the data (e.g. by abstracting point locations to being within a certain region, and by choosing one of a small finite set of relations rather than exact, but inaccurate relation over real valued data). By comparison with a conventional CV approach, where approximation may lead to inaccuracies, a qualitative approach will typically be indefinite but accurate. This approach can deal with some modes of input error (e.g. additive noise) but may fail to deal with other error modes (e.g. missing data, erroneous extra data). The use of conceptual neighbourhoods in [4] can help deal with certain kinds of missing data by allowing interpolation of “missing” intermediate relations in an event sequence, or by using them to filter out noise. Conceptual neighbourhoods may also be used to help predict the next qualitative state (cf [31]).

4 A Logical Approach to Cognitive Vision

Computer Vision falls into a class of problems where some sensor data, Ω , is acquired, and has to be interpreted relative to some already existing body of knowledge. Typically this body of knowledge falls into two categories: a very general, usually relatively domain independent knowledge base, Σ , and a more specific one, Φ , which may depend much more on the task(s) at hand. The problem is to explain the sensor data Ω given the prior knowledge. From a logical point of view, we can express this thus – what explanation Δ makes the following statement true:

$$\begin{array}{c} \downarrow \quad \downarrow \quad \downarrow \\ \Sigma, \Phi, \Delta \models \Omega \\ \downarrow \end{array} \quad (1)$$

The arrows above the formula indicate that these items are ‘inputs’, whilst the arrow below the formula indicates that Δ is the output – the abduced explanation.

This form of inference is called *abduction*. Shanahan [32,33] has applied this form of inference to the problem of abducting maps from robotic (non video) sensor data – see also [34]. More recently, he has also applied this approach to robotic vision [35] where he proposes to use abduction to formally explain all visual data either as a picture object or as noise, and preferring explanations with a higher “explanatory value”, i.e. which explain as little as possible as noise. The abduced explanations can then be used to feedback into the sensory action planning of the robot: it may initiate sensory actions to verify abduced hypotheses (e.g. by adjusting its noise thresholds, or even by attempting to touch or nudge a hypothesised object).

In the cognitive vision setting, we could view Σ as a background spatio-temporal theory (which might include, e.g. RCC-8), Φ as a set of possible behaviour patterns expressed in Σ (e.g. being stationary, bouncing up and down, descending,...), Ω is a qualitative abstraction of the spatio-temporal video data, and Δ is a set of behavioural instances which make the entailment true (i.e. explain the observations).

This is the core of our framework: the entire cognitive vision system is driven by the need to abduce explanations of sensor data given prior background knowledge.

A major issue is where does the background knowledge Σ and Φ come from? Very frequently, even in cognitive vision systems such as [3], this knowledge is explicitly programmed in by the human system builder. However this can be a tedious and often error prone process. Moreover, in a dynamic situation, or where the system is to be applied in different settings or contexts, different Σ and in particular Φ will be required, which adds greatly to the difficulty of acquiring such knowledge.

In fact, just about any computer vision system can probably be viewed in this way, whether it is explicitly logic based or not. What distinguishes our proposed approach is that:

- we will explicitly build our system around the entailment (1), and will use logical inference methods.
- the candidate hypotheses are expressed in a (largely) qualitative spatio-temporal representation language.
- we will, as far as possible, acquire Φ and possibly Σ too, automatically, as an *inductive* learning process from actual sensor data.

We argue that this has several advantages:

- The processes involved in interpreting visual data can be easily explained in terms of standard logical reasoning steps.
- Using a *qualitative* spatio-temporal representation allows easier generalisation of classes of different behaviours, guards against propagation of errors, and may facilitate the cognitive comprehension of spatio-temporal knowledge.
- Different variants of the architecture can be viewed as variants in control of the inferential process, different representation languages, different background knowledge bases, whilst maintaining a common architecture.
- The ability to learn the background knowledge makes the system much more robust and easier to field in different situations and domains.
- The use of logic allows the explicit manipulation of alternative hypotheses and the general statement of background knowledge. It also allows partial and indefinite knowledge to be represented. Equally importantly, such knowledge might be reused in other tasks associated with an artificial agent such as planning or map building.

5 Our Existing Implementations

In [4], we have shown how qualitative spatio-temporal models of events in traffic scenes (e.g. following, overtaking) can be learnt. Using an existing tracking program which generates labelled contours for objects in every frame, the view from a fixed camera is partitioned into semantically relevant regions based on the paths followed by moving objects. The paths are indexed with temporal information so objects moving along the same path at different speeds can be distinguished. Using a notion of proximity based on the speed of the moving objects and a description of the relationship between close objects using QSR calculi for relative direction and relative orientation with respect to the path being travelled, event models describing the behaviour of pairs of moving objects can be built, again using statistical methods. The system has been tested on a traffic domain and learns various event models expressed in the qualitative calculus which represent human observable events, e.g. following and overtaking. The system can then be used to recognise subsequent selected event occurrences or unusual behaviours. Although not explicitly encoded as abductive reasoning, this recognition phase can be viewed in this way: at each time step various behaviours may be possibly applicable given the current observations and the system keeps track of all the different possible explanations of the data. The system actually has recorded statistical frequency data during the learning phase and could use this to rank order the hypotheses. In the actual implementation they are all equally ranked and kept as possible explanations until subsequent observations rule out particular behaviours.

In newer work [36] based on [37,38], we are also interested in automatically inferring models of object interactions that can be used to interpret observed behaviour within a scene. Low-level computer vision techniques together with an attentional control mechanism are used to identify interesting incidents or events that occur in the scene over long periods of time. A data driven approach has been taken in order to automatically infer discrete and abstract representations (*symbols*) of primitive object interactions; this can be viewed as learning the basic qualitative spatial language of Σ . These symbols are then used as an alphabet to infer the high level structure of typical interactive behaviour using variable length Markov models (VLMMs) [39,40]. VLMMs deal with a class of random processes in which the memory length varies, in contrast to an n th-order Markov model for which the memory length is fixed. They have been previously used on the data compression [41,42] and language modelling domains [39,40,43] and recently, they have been successfully introduced in the computer vision domain for automatically inferring stochastic models of the high level structure of complex and semantically rich human activities [37,38]. It should be noted that although we are currently concentrating on applications within the traffic domain, our method is applicable to the general automatic surveillance task since it does not assume *a priori* knowledge of a specific domain. It is also worth pointing out explicitly that the use of probabilistic reasoning here contrasts with conventional low level use of probabilities – here the Markov model concerns high level semantic notions.

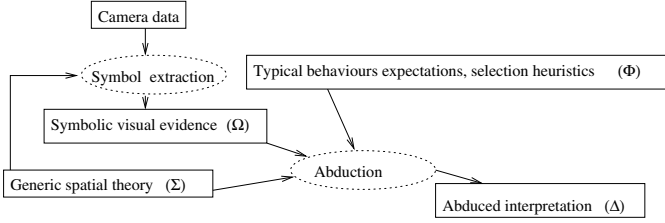


Fig. 2. System Overview: abducting interpretations

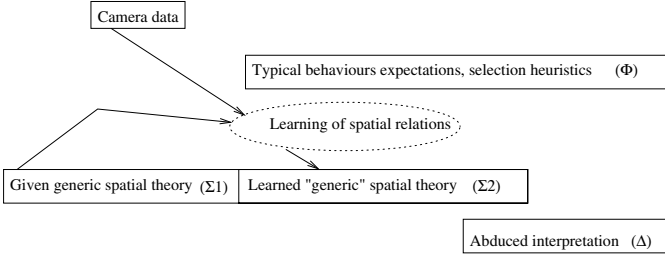


Fig. 3. System Overview: learning spatial relations

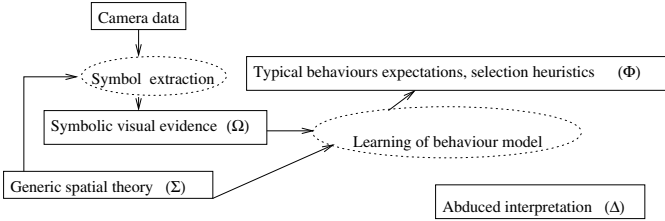


Fig. 4. System Overview: learning behavioural models

Figures 2, 3 and 4 give an overview of the system including the learning element of the architecture whereby the typical behaviour patterns are learned, stored and used to drive interpretation. These are the behavioural patterns, Φ , referred to in section 4 above. A real time computer vision system [44] detects and tracks moving objects within a scene. For each moving object scene feature descriptors are extracted that describe its relative motion and spatial relationship to all moving objects that fall within its attentional window (see [36] for details). These scene feature descriptors are invariant of the absolute position and direction of the interacting objects within a scene and are termed Ω in the figures.

Figure 2 shows the operation of the system once learning has taken place. Figure 3 shows how learning of part of the generic spatial theory Σ might take place. Figure 5 illustrates learnt primitive interactions for a traffic domain example application[36]. These can be viewed as a qualitative discretisation of the

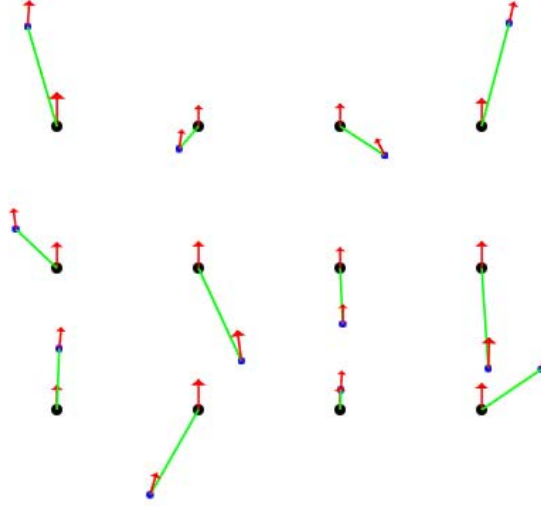


Fig. 5. Learnt primitive interactions – traffic domain example. The two dots represent pairs of close vehicles (distinguished by the size of the dot). The arrows show their direction of movement and the connecting vector their relative orientation. These patterns represent typical “midpoints” as result of clustering the input data into n different conceptual “regions”. Note how the 12 relations naturally cluster into virtually symmetric pairs, e.g. the middle two prototypes on the first line.

continuous relational space. Whereas in a conventional QSR representation, the discretisation would be manually preassigned, we are able to learn a representation which maximises the discernability given a granularity (i.e. the number of relations desired). The system can currently be used to recognize typical interactive behaviour within the traffic domain and identify atypical events. These learnt primitive interactions effectively form part of the background spatial theory, which is labelled Σ_2 in figure 3.

In figure 4 we show how behaviours might be learnt. In [36] we use a statistical learning framework [45] where discrete representations of interactive behaviours can be learned by modelling the probability distribution of the primitive interactions. VLMMs are then used to efficiently encode the sequences of these learned primitive patterns corresponding to observed interactive behaviour.

6 Incorporating QSR into the Heart of a Computer Vision System

As already noted, QSR has been used as a post-processing method on the output of a quantitative real-world analysis system (such as a vision system), e.g. [4]. In this section we will propose an alternative approach that puts QSR methods at the heart of a computer vision system. This will have the effect of constraining

the output of the system to be logically consistent (with respect to the QSR theory embodied in Σ and Φ). This is done by using low level CV algorithms (e.g. colour region segmentation algorithms) that draw no conclusions about the nature or structure of the data. The output of the low level process thus makes as few semantic inferences as possible. For example it will not embody an “object tracker” as that would presuppose the ability to recognise objects. What it does do, is at each time step (frame) to distinguish certain spatial elements, and assign certain qualitative properties to them (such as colour, texture and qualitative spatial relationships). The sequence of these outputs comprises Ω . The spatial elements in Ω thus become the primitive spatial elements (rather than the original pixels); like pixels they may be “mixed”, in the sense that they contain elements of different objects, but they will never be split apart, but the heterogeneity will be symbolically reasoned about.

The higher level reasoning component then comprises three principal mechanisms:

- A qualitative spatio-temporal reasoner which uses Σ and performs certain inferences such as checking consistency both statically and with respect to continuous motion for example.
- An abduction engine which uses Φ in conjunction with the other data to generate hypotheses, i.e. possible (partial) explanations.
- A technique for handling uncertainty: the behaviours in Φ may be probabilistic or have other metadata indicating their absolute or relative likelihood. This component should ensure that the most likely explanation is chosen.

The objective of our proposed system is to explain a complete observation (or set of observations) rather than a subset of the objects within the scene as is traditionally the task. This removes the logical distinction between objects and ‘background’ (the rest of the scene). This has the practical disadvantage of causing the computational cost of any QSR method to be exponential in the granularity of the problem.

To cope with the computational explosion caused by this explicit use of symbolic spatial reasoning, we borrow two techniques used in human perception: attention and multi-resolution/scale processing. A good attention mechanism commonly used in CV is motion. This is highly suitable for online reasoning as moving areas contain more spatio-temporal object information than static areas and as such require processing at a higher rate/resolution.

6.1 Scene Understanding as Scene Explanation

We wish to represent a single observation of a scene as resulting from a number of objects (with no concept of background). In the real world, scene description is not this simple as objects exist in a conceptual hierarchy. Figure 6 gives an example of such a hierarchy.

Figure 6 illustrates that many conceptual objects are in fact composite objects constructed by the combination of simpler objects or composite objects.

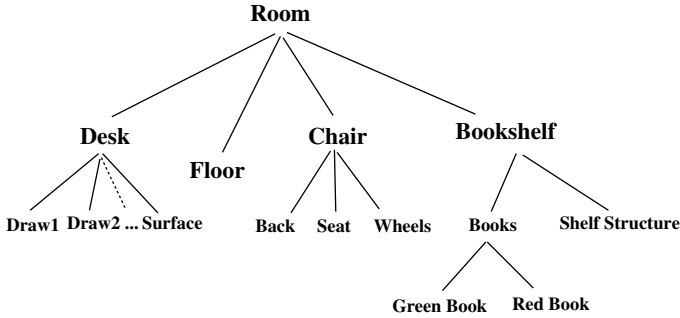


Fig. 6. Conceptual Hierarchy of Objects for a Room

This conceptual hierarchy sits on a single level observational hierarchy in which base level conceptual objects are divided into non-semantic observational regions. An example of this is given in figure 7(a).

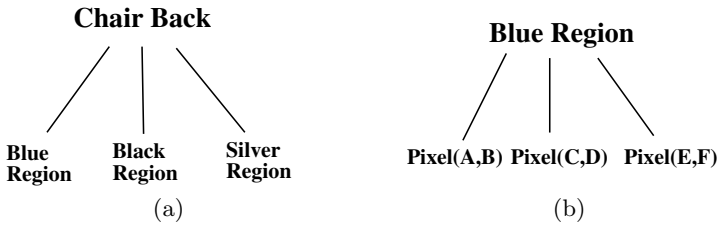


Fig. 7. (a) Observational Hierarchy of Chair Object (b) Sensory Hierarchy of a Region: a particular region ("Blue Region") is composed of a number of pixels at particular x, y coordinates.

From the point of view of a CV system (or human vision system) the conceptual hierarchy sits on a sensory hierarchy with atomic sensory components (pixels in the case of a CV system). An example of this is given in figure 7(b).

6.2 Automatic Building of Object Hierarchies

The purpose of any scene understanding system is to automatically build complete or (more usually) partial object hierarchies of the nature of those described in the previous section from the bottom up. Alternatively, *a priori* information may be used in a 'hypothesise and test way' to build object hierarchies from the top down. The bottom up approach is essentially limited to building sensory and observational hierarchies as no *a priori* conceptual information is available. The top down approach can build complete hierarchies; however an *a priori* model for any object or composite object that may occur in a scene must be available. This is a problem if we wish to build complete hierarchies of complex scenes.

In many real world CV systems a combination of the top down and bottom up approaches is used [44,46]; however the interface of these two approaches is often *ad hoc*. This can lead to errors and logical inconsistencies in the final scene analysis. We propose to use QSR to interface low level (bottom up) approaches with high level (top down) methods in such way that low level logical inconsistencies do not occur in the high level interpretation. An example of this would be to use continuity networks such as the one in figure 1 to filter out low level spatio-temporal data which are not continuous with respect to this diagram (e.g. if disconnected regions are immediately afterwards partially overlapping). A refinement to this approach is to use continuity networks which are specialised to the kinds of objects involved. In [47,30] we distinguish various weaker notions of continuity which may be appropriate for certain kinds of objects and correspondingly weaker conceptual neighbourhood diagrams. If the vision system can recognise the types of the objects involved, then the notion of continuity can be correspondingly specialised.

We propose to use bottom up CV to build sensory hierarchies that describe the entire image and use abduction to generate (over time) a set of logically consistent hypotheses for the complete scene description hierarchy. Higher level (top down) CV methods incorporating *a priori* knowledge would then be used to validate and rank these hypothesis and assign semantic labels. Objects in a scene for which no *a priori* information exists would remain unvalidated and unlabelled; however this would be explicitly flagged by the system and could be used as the basis of a novel object learning system.

Past time hypotheses may be declared invalid given subsequent observations and deleted. For reasons of computational tractability it may be necessary to only consider a small window into the past when evaluating the validity of past hypotheses.

Our approach to abduction is outlined in more detail in [48]. In essence, the problem is to determine, given a background spatio-temporal theory, a set of typical patterns of behaviour and a set of qualitative spatio-temporal observations, what actual objects and behaviours could explain the observations.

6.3 Reasoning over Time and Hypothesis Verification

QSR and abduction will generate scene hypotheses for the complete scene description hierarchy at the current timestep and at previous timesteps and the validity relationships between these over time. This is illustrated in figure 8. As can be seen, in general there will be more than one possible explanation abduced and a way of rank ordering the various hypotheses offered as explanations will be needed too. In [48] we give some logic based techniques whereby a preferred hypothesis (or set of preferred hypotheses) might be selected. In the dynamic case we are considering here, we will want to carry forward multiple hypotheses from one frame to the next and use information gained from future frames as well as statistically and *a priori* knowledge based heuristics to choose a single preferred hypothesis when required.

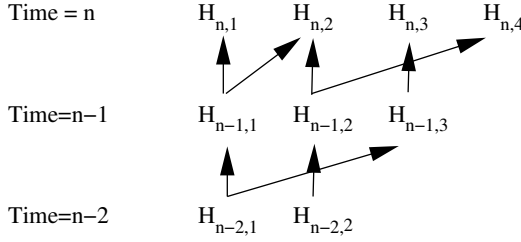


Fig. 8. Hypotheses are Generated over Time by QSR and Abduction

Using *a priori* knowledge, a hypothesis may be assigned low probability at the current timestep and this can be propagated as a future hypothesis. *A priori* knowledge can exist on many levels from detailed object models (e.g. 3D shape or texture models) to more general object class models (e.g walls are static and of homogeneous colour and texture). In a hypothesis verification scenario these different types of *a priori* knowledge may be combined so as to verify the complete hypotheses.

Methods used to encapsulate typical (statistical) *a priori* spatial or temporal object information in the scheme described so far must be able to take as their input an object (or composite object) description hierarchy (or sequence of these). These hierarchies may be thought of as a list of atomic elements containing their properties (in addition to structural information which may be ignored if necessarily). These atomic elements may be at the sensory level (pixels) or the observational level (homogeneous regions). This does not fit well with the traditional N-dimensional object models described in section 2. These models must be adapted or replaced to fit in with our proposed variable length list object descriptions. How such methods are to be formulated is the subject of current work and beyond the scope of this paper; however it should be noted that comparisons between model and observation are on an object-object basis rather than an object-scene basis as with many traditional CV methods. This allows observation to model matching in addition to model to observation matching.

7 Future Work

We are planning a wide variety of future work in order to flesh out and validate our proposed architecture, not only within the traffic domain but also in another domain, for example a “kitchen or table top scenario”. There is theoretical work to do as well as actually implementing a system conforming to the ideas presented here.

In particular, further research in qualitative spatial and spatio-temporal representation and reasoning will be required. Much work has concentrated on topological and mereotopological calculi to date as indicated in [12]. New calculi such as the occlusion calculus [49], are being specifically developed for cognitive vision

– the occlusion calculus being specifically targeted at the problem of reasoning about the topology of occluded regions.

However, in order to model and distinguish between different kinds of objects that might be found in visual scenes, not only will other aspects, such as orientation and size become important, but more particularly, the notion of qualitative shape will need to be further developed so as to be useful for cognitive vision. There is already some work on orientation (e.g. [18,50]), though this is largely point based, rather than region based, which may prove to be more useful. There is relatively little work still on qualitative shape representations. Existing work includes boundary based approaches [51,52], representation through elongation and symmetric aspects [53], and the use of a convex hull primitive (which essentially gives an affine geometry [54]). However the utility of these approaches applied to cognitive vision has not been tested to any great extent.

It is worth pointing out that although in general it is very hard to represent shape in a qualitative way since very small changes in shape may lead to very different functionality (e.g. consider interlocking gears), in cognitive vision the task is not so much to reason about kinematics (or similar predictive/analytical tasks which require detailed shape knowledge) but rather simply to categorise object shape in order to classify and recognise different kinds of objects. Arguably this task will be easier, but this has not been particularly investigated.

A vital aspect of a cognitive vision architecture is the ability to represent and reason about extended event sequences. Although VLMMs have been successfully applied in computer vision in order to represent long-term behaviours, criticisms may be made of them that they have no semantics in themselves. The use of a qualitative spatio-temporal representation within a logical framework holds out the promise of a formally defined semantics, and a richer vocabulary to describe extended behaviours. Moreover, the notion of continuity present in the conceptual neighbourhoods of a qualitative spatial calculus may be used to help constrain the learning and interpretation of event sequences. However more research is required in order to validate this hypothesis and to develop a qualitative spatio-temporal theory that is well adapted to the demands of cognitive vision.

Finding control architectures to moderate the inferential mechanism to process data efficiently is a key research question, for example to develop a controllable attentional mechanism. Also on our priority list is to develop further techniques for learning the background knowledge Σ and particularly Φ .

We also plan to consider other input features apart from orientation, relative direction of motion and distance, e.g. acceleration and shape. Another consideration is to consider how non interactive behaviours may be represented and reasoned about (at present the feature descriptions are always between pairs of moving objects). Finally, we hope to integrate our approaches with some of those in our partners in the CogVis project IST-2000-29375 and to take account of other work such as the approach to learning from low level data of [55].

Finally, we plan to consider the evaluation of our cognitive vision system: under what circumstances would we say that we have succeeded? Clearly, we

can inspect the internal architecture of the system and the extent to which it has high level representations, the extent to which it can learn and meet the other considerations mentioned in the introduction. A further criterium would be to evaluate with respect to human visual cognition; for example Tversky [56,57] has investigated the perception of the event structure of a video sequence by human subjects – can we produce a cognitive vision system which can infer a similar structure?

Acknowledgements

The support of the EPSRC under grant GR/M56807 and the EU under IST-2000-29375 is gratefully acknowledged.

References

1. Yanai, K., Deguchi, K.: Recognition of indoor images employing qualitative model fitting and supporting relation between objects. In Sanfeliu, A., Villanueva, J., Vanrell, M., Alquezar, R., Eklundh, J.O., Aloimonos, Y., eds.: Proceedings 15th International Conference on Pattern Recognition. Volume 1., Barcelona, Spain, IEEE Press (2000) 964–967
2. Howarth, R.: Interpreting a dynamic and uncertain world: High-level vision. *Artificial Intelligence Review* **9** (1995) 37–63
3. Buxton, H., Howarth, R.: Spatial and temporal reasoning in the generation of dynamic scene descriptions. In Rodríguez, R.V., ed.: Proceedings on Spatial and Temporal Reasoning, Montréal, Canada, IJCAI-95 Workshop (1995) 107–115
4. Fernyhough, J., Cohn, A., Hogg, D.: Constructing qualitative event models automatically from video input. *Image and Vision Computing* **18** (2000) 81–103
5. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Training models of shape from sets of examples. In: Proc. British Machine Vision Conference. (1992) 9–18
6. Baumberg, A., Hogg, D.: Learning flexible models from image sequences. In: European Conference on Computer Vision, Springer Verlag (1994) 299–308
7. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. In: Proc. First International Conference on Computer Vision. (1989) 259–268
8. Blake, A., Curwen, R., Zisserman, A.: A framework for spatiotemporal control in the tracking of visual contours. *International Journal of Computer Vision* **11** (1993) 127–145
9. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* **3** (1991) 71–86
10. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77** (1989) 257–286
11. Starner, T., Pentland, A.: Real-time american sign language recognition from video using hidden markov models. In: Int. Symposium on Computer Vision. (1995)
12. Cohn, A.G., Hazarika, S.M.: Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae* **46** (2001) 1–29
13. Cohn, A.G., Bennett, B., Gooday, J., Gotts, N.: RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica* **1** (1997) 275–316
14. Egenhofer, M., Franzosa, R.: Point-set topological spatial relations. *International Journal of Geographical Information Systems* **5** (1991) 161–174

15. Cohn, A.G., Gotts, N.M.: Representing spatial vagueness: a mereological approach. In L C Aiello, J.D., Shapiro, S., eds.: Proceedings of the 5th conference on principles of knowledge representation and reasoning (KR-96), Morgan Kaufmann (1996) 230–241
16. Clementini, E., Di Felice, P.: Approximate topological relations. *International Journal of Approximate Reasoning* **16** (1997) 173–204
17. Schlieder, C.: Reasoning about ordering. In A Frank, W.K., ed.: *Spatial Information Theory: a theoretical basis for GIS*. Number 988 in Lecture Notes in Computer Science, Berlin, Springer Verlag (1995) 341–349
18. Isli, A., Cohn, A.: A new approach to cyclic ordering of 2d orientations using ternary relation algebras. *Artificial Intelligence* **122** (2000) 137–187
19. Frank, A.U.: Qualitative spatial reasoning about distance and directions in geographic space. *Journal of Visual Languages and Computing* **3** (1992) 343–373
20. Meathrel, R.C., Galton, A.P.: A heirarchy of boundary-based shape descriptors. In Nebel, B., ed.: *Proc. 17th IJCAI*, Morgan Kaufmann (2001) 1359 – 1364
21. Bennett, B., Cohn, A.G., Torrini, P., Hazarika, S.M.: Describing rigid body motions in a qualitative theory of spatial regions. In Kautz, H.A., Porter, B., eds.: *Proceedings of AAAI-2000*. (2000) 503–509
22. Cristani, M., Cohn, A., Bennett, B.: Spatial locations via morpho-mereology. In: *Proc. KR'2000*, Morgan Kaufmann (2000)
23. Cohn, A.G., Bennett, B., Gooday, J., Gotts, N.: Representing and reasoning with qualitative spatial relations about regions. In Stock, O., ed.: *Temporal and spatial reasoning*, Kluwer (1997)
24. Wolter, F., Zakharyashev, M.: Spatio-temporal representation and reasoning based on RCC-8. In: *Proceedings of the seventh Conference on Principles of Knowledge Representation and Reasoning*, Morgan Kaufman (2000) 3–14
25. Wolter, F., Zakharyashev, M.: Qualitative spatio-temporal representation and reasoning: a computational perspective. In: *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann (To appear)
26. Bennett, B., Cohn, A., Wolter, F., Zakharyashev, M.: Multi-dimensional modal logic as a framework for spatio-temporal reasoning. *Applied Intelligence* (2002) To appear.
27. Hayes, P.J.: Naive physics I: Ontology for liquids. In Hobbs, J.R., Moore, B., eds.: *Formal Theories of the Commonsense World*. Ablex (1985) 71–89
28. Muller, P.: A qualitative theory of motion based on spatio-temporal primitives. In Cohn, A.G., Schubert, L.K., Shapiro, S., eds.: *Principles of Knowledge Representation and Reasoning: Proceedings of the 6th International Conference (KR-98)*, Morgan Kaufman (1998) 131–141
29. Muller, P.: Space-time as a primitive for space and motion. In Guarino, N., ed.: *Formal ontology in information systems: Proceedings of the 1st international conference (FOIS-98)*. Volume 46 of *Frontiers in Artificial Intelligence and Applications*., Trento, Italy, Ios Press (1998) 63–76
30. Hazarika, S.M., Cohn, A.G.: Qualitative spatio-temporal continuity. In Montello, D.R., ed.: *Spatial Information Theory: Foundations of Geographic Information Science; Proceedings of COSIT'01*. Volume 2205 of *LNCS*., Morro Bay, CA, Springer (2001) 92–107
31. Cui, Z., Cohn, A.G., Randell, D.A.: Qualitative simulation based on a logical formalism of space and time. In: *Proceedings of AAAI-92*, Menlo Park, California, AAAI Press (1992) 679–684
32. Shanahan, M.: Noise, non-determinism and spatial uncertainty. In: *Proceedings of AAAI-97*. (1997) 153–158

33. Shanahan, M.: A logical account of the common sense informatic situation for a mobile robot. *Electronic Transactions on Artificial Intelligence* (1999)
34. Remolina, E., Kuipers, B.: A logical account of causal and topological maps. In: *Proceedings of Seventeenth International Conference on Artificial Intelligence (IJCAI-01)*. Volume I, Seattle, Washington, USA (2001) 5–11
35. Shanahan, M.: A logical account of perception incorporating feedback and expectation. In: *Proc. 8th Int. Conf. on Knowledge Representation and Reasoning*, San Mateo, Morgan Kaufmann (2002)
36. Galata, A., Cohn, A.G., Magee, D., Hogg, D.: Modelling interaction using learnt qualitative spatio-temporal relations and variable length markov models. In: *Proc. European Conference on AI (ECAI)*. (2002)
37. Galata, A., Johnson, N., Hogg, D.: Learning behaviour models of human activities. In: *British Machine Vision Conference*, BMVC'99. (1999)
38. Galata, A., Johnson, N., Hogg, D.: Learning Variable Length Markov Models of Behaviour. *Computer Vision and Image Understanding (CVIU) Journal* **81** (2001) 398–413
39. Ron, D., Singer, S., Tishby, N.: The Power of Amnesia. In: *Advances in Neural Information Processing Systems*. Volume 6. Morgan Kauffmann (1994) 176–183
40. Guyon, I., Pereira, F.: Design of a Linguistic Postprocessor using Variable Memory Length Markov Models. In: *International Conference on Document Analysis and Recognition*. (1995) 454–457
41. Cormack, G., Horspool, R.: Data Compression using Dynamic Markov Modelling. *Computer Journal* **30** (1987) 541–550
42. Bell, T., Cleary, J., Witten, I.: *Text Compression*. Prentice Hall (1990)
43. Hu, J., Turin, W., Brown, M.: Language Modelling using Stochastic Automata with Variable Length Contexts. *Computer Speech and Language* **11** (1997) 1–16
44. Magee, D.: Tracking multiple vehicles using foreground, background and motion models. In: *Proc. ECCV Workshop on Statistical Methods in Video Processing*. (2002)
45. Johnson, N., Hogg, D.: Learning the Distribution of Object Trajectories for Event Recognition. *Image and Vision Computing* **14** (1996) 609–615
46. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfnder: Real-time tracking of the human body. *IEEE Transactions on PAMI* **19(7)** (1997) 780–785
47. Cohn, A.G., Hazarika, S.M.: Continuous transitions in mereotopology. In: *Commonsense-2001: 5th Symposium on Logical Formalizations of Commonsense Reasoning*. (2001)
48. Hazarika, S.M., Cohn, A.G.: Abducing qualitative spatio-temporal histories from partial observations. In: *Proc. 8th Int. Conf. on Knowledge Representation and Reasoning*, San Mateo, Morgan Kaufmann (2002)
49. Randell, D., Witkowski, M., Shanahan, M.: From images to bodies: Modelling and exploiting spatial occlusion and motion parallax. In: *Proc. IJCAI*, Morgan Kaufmann (2001)
50. Freksa, C.: Using orientation information for qualitative spatial reasoning. In: Frank, A.U., Campari, I., Formentini, U., eds.: *Proc. Int. Conf. on Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, Berlin, Springer-verlag (1992)
51. Meathrel, R.C., Galton, A.: A hierarchy of boundary-based shape descriptors. In: *Proc. IJCAI*. (2001) 1359–1364

52. Jungert, E.: Symbolic spatial reasoning on object shapes for qualitative matching. In Frank, A.U., Campari, L., eds.: *Spatial Information Theory: A Theoretical Basis for GIS*. Lecture Notes in Computer Science No. 716, COSIT'93, Springer-Verlag (1993) 444–462
53. Clementini, E., Di Felice, P.: A global framework for qualitative shape description. *Geoinformatica* **1** (1997) 1–17
54. Davis, E., Gotts, N.M., Cohn, A.G.: Constraint networks of topological relations and convexity. *Constraints* **4** (1999) 241–280
55. Kaelbling, L.P., Oates, T., Hernandez, N., Finney, S.: Learning in worlds with objects. In Cohen, P.R., Oates, T., eds.: *Learning Grounded Representations*. Number Technical Report SS-01-05, AAAI Press (2001) 31–36
56. Zacks, J., Tversky, B., Iyer, G.: Perceiving, remembering and communicating structure in events. *Journal of Experimental Psychology: General* **136** (2001) 29–58
57. Zacks, J., Tversky, B.: Event structure in perception and conception. *Psychological Bulletin* **127** (2001) 3–21

How Similarity Shapes Diagrams

Merideth Gattis

Department of Psychology, University of Sheffield, Western Bank, Sheffield S10 2TP, UK
GattisM@cardiff.ac.uk

Abstract. Most diagrams communicate effectively despite the fact that diagrams as a group have a minimum of conventions and a high tolerance for novelty. This paper proposes that the diversity and felicity of diagrammatic representation is based on three kinds of similarity between semantic propositions and spatial representations that allow people to interpret diagrams consistently with a minimum of effort and training. *Iconicity* is similarity of physical appearance, *polarity* is similarity in the positive and negative structure of dimensions, and *relational similarity* aligns structures so that elements correspond to elements, relations correspond to relations, and so on. In diagrammatic reasoning detected similarities are used to create correspondences between the visual characteristics of a diagram and its semantic meaning, and those correspondences are in turn used to make inferences about unknown or underspecified meanings.

1 Diagrams Represent Many Kinds of Relations in Many Kinds of Ways

Diagrams represent many kinds of relations – some spatial, some nonspatial, and some a mix between the two. Diagrams have been used to record activities, ownership, and places (see Tversky, 2001 for a review). Diagrams communicate what has happened in the past, what someone has in mind at the moment, which activities are or are not allowed in a particular place, and what may be expected to happen along a particular stretch of the road. Diagrams sometimes include text, sometimes include conventionalized visual symbols, and sometimes contain novel or unconventional representations. Thus as a group, diagrams seem to function like one of those clubs that lets nearly everyone inside. Perhaps the only necessary conditions for diagrams are 1) that they convey meaning, and 2) that they do so visuospatially.

Unrestricted membership and lack of conventions may be a fine way to run a social club, but in a communication system it normally leads to misunderstanding and confusion. One very special characteristic of diagrams is that most diagrams communicate effectively despite the modicum of conventions and the high tolerance for novelty among diagrams as a group. One indicator that diagrams do communicate effectively is that diagrams are often the default communication device for multilingual contexts in which people are likely to need to know something or say something, but do not share a common language that would enable them to do so. Airports, highways, and ports are full of diagrams that communicate things like “This is where you buy a ticket,” “This is where you change money,” and “Do not stand on

this ledge.” Diagrams and graphs play a similar role in scientific conferences, journals, and textbooks, effectively communicating more abstract relations such as “These two variables interacted,” “Either x holds or y holds but both cannot hold simultaneously,” and “Mechanism A is hypothesized to feed outputs to Mechanism B,” even when shared language is limited. Diagrams are also effective in helping people to reason about complex relations such as double disjunctions (Bauer & Johnson-Laird, 1993).

2 The Role of Similarity in Diagrams

In this paper I would like to argue that we are able to create and interpret diagrams that represent many kinds of relations in many kinds of ways because we are able to detect many kinds of similarity between semantic propositions and visuospatial representations. We then use detected similarities to create correspondences between the visual presentation of a diagram and its semantic meaning. Finally, in a process similar to analogical reasoning, we use those correspondences to make inferences about unknown or underspecified meanings.

Similarity refers to properties shared between two or more concepts, ideas, or representations (Tversky, 1977). Psychologists have known for quite some time that similarity influences perception. For instance, we tend to perceive similar objects as belonging together, as shown in Figure 1 (Wertheimer, 1923). Similarity also influences many aspects of thought, including analogical problem solving, category assignment, decision making, judgments of the likelihood of some event, and learning about science in the classroom (Vosniadou & Ortony, 1989). The fact that similarity influences both perception and thought suggests that similarity may play a particularly important role in the interpretation and use of visual representations such as diagrams, graphs, and maps.



Fig. 1. Because perceptual similarity influences visual grouping, people perceive the dots in the figure on the left as being organized into vertical columns and the dots in the figure on the right as being organized into horizontal rows, even though the dots are equally spaced in both diagrams.

Representation does not, however, always rely on similarity. For instance, languages rely for the most part on arbitrary representation. Similarity does not play an important role in which phonemes and morphemes mean “dog,” “eat,” and “world,” with the result that the statement in English “It’s a dog eat dog world” is

indecipherable to someone who doesn't understand English. A simple proof that at least some diagrams do rely on similarity for representation can be seen in a very simple diagram of the same statement (see Figure 2).

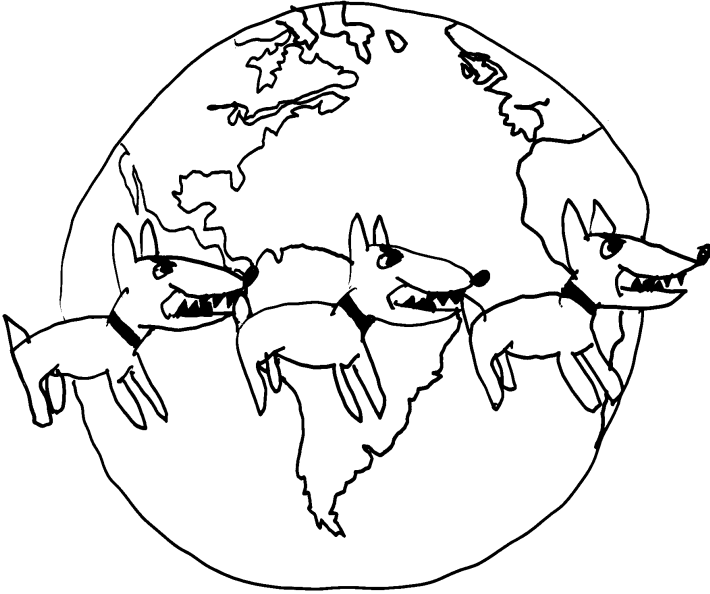


Fig. 2. Interpreting an iconic representation of the saying “It’s a dog eat dog world,” requires less specific knowledge than interpreting the same linguistic utterance because the elements in the diagram are physically similar to the things they represent (in a rough way).

Diagrams vary widely, however, and whether similarity plays an important role in the creation and interpretation of all diagrams is an open question. In this section of the paper, I will briefly review evidence for three types of similarity that may influence diagrammatic representation: iconicity, polarity, and relational structure (Gattis, 2001a; 2002). These three forms of similarity vary in the level of resemblance, and as a result in the level of meaning represented.

2.1 Iconicity

The most familiar form of similarity exhibited in diagrams is when diagrams contain elements that physically resemble the things they represent. Representation via physical similarity is known as iconicity, and has been discussed and studied by semioticians, linguists, and psychologists for years (Bertin, 1983; Fromkin & Rodman, 1998; Peirce, 1903/1960; Tversky, 1995). Figure 2, a diagram representing the saying “It’s a dog eat dog world,” illustrates the fact that iconicity does not require complete physical resemblance, but only some partial resemblance. As a result, iconicity is a flexible representational tool. A related advantage of iconicity is that it usually leads to easily interpretable representations: think again of the difference between knowing what the morpheme “dog” means, versus being able to recognize a

picture of a dog. This thought experiment illustrates what more controlled experiments have confirmed, that although the shared meanings conveyed through iconicity may to some extent be culturally determined, little or no specific training is required to understand iconic representations (Huer, 2000; Koul & Lloyd, 1998; Morford, 1996).

One clear exception to the advantages of iconicity is when representations attempt to convey complex relations or concepts. A friend once told me about a highway sign that he studied for a long time before finally concluding that the sign was attempting to convey iconically the complex relational concept that dogs might relieve themselves in the countryside, but humans should use the facilities provided for them by the highway services. This humorous story illustrates an important point about iconicity, namely that iconicity is a limited representational tool because some meanings are difficult to represent iconically. This is particularly true for representations of relations and abstract concepts. Signs instructing people where to conduct a particular activity, for instance, frequently rely on representing the objects involved in the activity (e.g. some money, for money changing, or a cup, for drinking). Concepts such as peace, love, and egalitarianism are less amenable to iconicity, and the same is true for many relations.

Luckily, although iconicity is a ubiquitous tool in diagrammatic representation, it is not the only one. Some diagrams do not physically resemble the things they represent, yet people tend to interpret these diagrams in consistent ways, even when they do not have specific experience or training with the diagrams. The cause of this surprising consistency appears to be that we rely on more abstract forms of similarity to create correspondences between diagrams and their meanings. Representations of relations and concepts seem to rely on two abstract forms of similarity, both having to do with the semantic structure of concepts rather than physical resemblance. The first of these forms of similarity concerns polar structure, and the second concerns relational structure.

2.2 Polarity

Many of the abstract concepts about which we communicate can be described as polar dimensions. Studies from linguists and psycholinguists have demonstrated that the words describing physical dimensions such as amount, brightness, length, depth, size, temperature, and weight, and more abstract dimensions such as age, activity, generosity, and goodness have either a positive or negative “weight” (Gilpin & Allen, 1974; Hamilton & Deese, 1971). For each of these dimensions, one term is used to describe the entire dimension as well as a particular end of the dimension (e.g., more, long, generous, and good), and thus is more general, or positively weighted. A second term is used to describe just one end of the dimension and is understood with reference to the first term (e.g., less, short, stingy, bad). The second term is more specific, and therefore is described as being negatively weighted.

The positive and negative structure of dimensions is not exclusively a property of language, but of perception as well. Psychophysical studies in the 1950’s and 1960’s established that many perceptual dimensions have a polar structure (Stevens, 1975). Dimensions such as loudness, brightness, hardness, and roughness each have one end which is the primary attribute of that dimension. When asked to match stimuli with varying perceptual properties, people tend to match primary attributes of different

perceptual dimensions. As in language, in perception, up, loud, rough, and hard are positively weighted. These cross-modal matching experiments from psychophysicists and developmental psychologists have demonstrated that polarity is an important form of similarity across different dimensions and modalities. The obvious advantage of this is that because polarity is an abstract and cross-modal form of similarity, it may be a basis for creating and interpreting diagrammatic representations.

Because polarity is a property of both perception and language, it makes sense that it would play an important role in diagrammatic representation. Recent studies confirm this hypothesis (Gattis, 2001b, c). In several studies, children and adults were given a cross-modal matching task that resembled previous studies of psychophysicists and developmental psychologists. The crucial difference was that instead of being asked to match two physical stimuli, participants were asked to match two function lines in a Cartesian graph with two animals with contrasting traits along dimensions such as size, loudness, and achromatic hue (see Figure 3). Adults

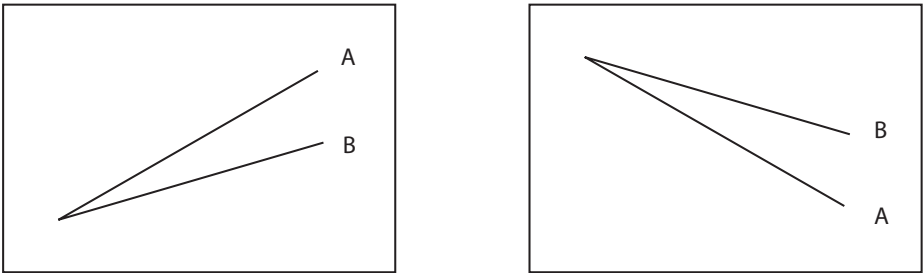


Fig. 3. Interpreting diagrams sometimes involves establishing correspondences between the polar structure of perceptual dimensions in a graph, such as height, length, and slope, and the polar structure of linguistic dimensions being represented, such as size or loudness. Children and adults shown one of the above diagrams identified the lines marked A as representing a loud dog or a big bear, and the lines marked B as representing a quiet dog or a little bear.

and children as young as 3 years old interpreted the diagrams in a way that created a correspondence between the perceptual polarity of the diagrams and the linguistic polarity of the traits in question. For example, when told a story about dogs, one of whom is loud and one of whom is quiet, and asked which line stands for which dog, both adults and children as young as 3 years identified the top line in the left diagram and the bottom line in the right diagram as “the loud dog.” Similarly, children and adults identified the bottom line in the left diagram and the top line in the right diagram as “the quiet dog.” This pattern of judgments suggests that even very young children were sensitive to the perceptual polarity of the slope of the lines, and the linguistic polarity of the dimensions along which the animals varied, and used that polarity to establish cross-modal correspondences between the diagram and its meaning.

2.3 Relational Structure

A third form of similarity influences the interpretation of diagrams because it helps the reader establish correspondences based not on physical resemblance or on

correspondences between dimensions, but on similarities of the relational structure as a whole. Similarities of relational structure lead to a diagrammatic mapping of individual elements of meaning, as well as a mapping of relations between elements and relations between dimensions (Gattis, 2001a; 2002). Many studies of analogical reasoning indicate that when comparing two problems or two concepts, adults tend to align the problems in such a way as to map elements of one concept to elements of another, relations between elements in one concept to relations between elements in another, and so on. Sensitivity to relational structure emerges in early childhood. By the age of 4, children are sensitive to the relational structures of perceptual analogy tasks, and by the age of 6 children will align the relational structures to choose “matches” to complete a perceptual analogy (Kotovsky & Gentner, 1996).

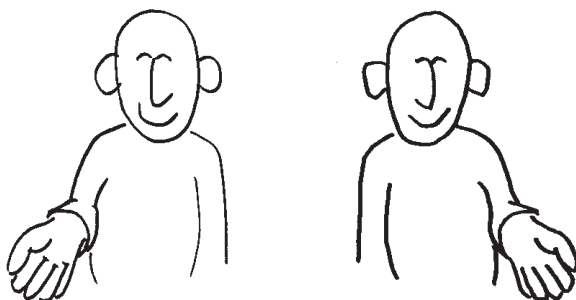


Fig. 4. Interpreting diagrams sometimes involves establishing correspondences between semantic elements and physical elements, and between semantic relations and physical relations. In several experiments investigating how relational correspondences are established during diagram interpretation, people were presented with the diagrams shown above paired with a specific meaning, such as “This hand means Mouse” and “This hand means Bear.” Figures 5 and 6 explain how the experiments continued.

Recent experiments in my lab have investigated the role that similarity of very simple relational structures – elements and relations between elements – plays in interpretation of diagrams (Gattis, 2001d). In these experiments drawings of a man grasping his ears are paired with meaningful statements, and then participants are asked to judge the meaning of similar but new diagrams. In one experiment, for instance, participants first saw two drawings (see Figure 4) of a man extending his right and left hand, and each drawing was paired with a statement assigning a specific meaning to each hand, such as “This hand means Mouse,” and “This hand means Bear.” Participants then saw two new drawings (see Figure 5) of the man touching his right ear and his left ear, each time with his right hand. Each drawing was paired with a statement about the animal represented by that hand performing involved in some action with another animal, and the two statements differed either in the subject, the object, or the relation between the subject and object (i.e. “Monkey bites Mouse” and “Elephant bites Mouse,” or “Mouse bites Monkey” and “Mouse bites Elephant,” or “Mouse bites Monkey” and “Mouse visits Monkey.” In all of these cases, an ambiguity existed between the diagram and the statements: the varying meaning could be assigned to the varying elements involved (i.e. the ears) or to the varying relations involved (i.e. the relation of the arm to the body).

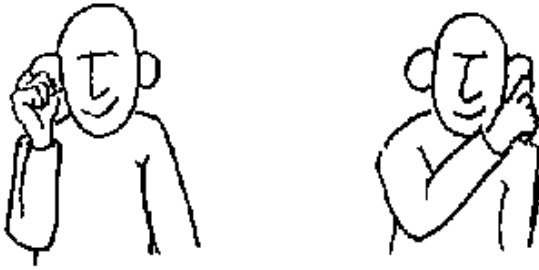


Fig. 5. After the diagrams in Figure 4, people were presented with two diagrams each paired with a meaningful statement. The two statements were identical except for an element that varied (i.e. the subject or object of an action) or a relation that varied (i.e. the action itself). Examples of these statements are “Mouse bites Monkey” and “Mouse bites Elephant,” or “Mouse bites Monkey” and “Mouse visits Monkey.” The varying meaning may be assigned to the varying elements involved (i.e. the ears) or the varying relations involved (i.e. the relation of the arm to the body).

People were then asked to make a judgment about the meaning of new diagrams showing the same character making similar gestures with his other hand (see Figure 6). People chose between two possible meanings for each new diagram, and the two possibilities differed in the same way as the previous phase (i.e. the subject, the object, or the relation). For example, the choices might have been “Monkey bites Bear” or “Elephant bites Bear” for the subject-varying condition, “Bear bites Monkey” or “Bear bites Elephant” for the object-varying condition, or “Bear bites Monkey” or “Bear visits Monkey” for the relation-varying condition.

The judgment was intended to probe how people resolved the ambiguous assignment of meaning within each diagram. By comparing the meaning chosen for a particular diagram with the two previously assigned meanings, it was possible to diagnose whether meaning had been assigned to the varying elements involved (i.e. the ears) or to the varying relations involved (i.e. the relation of the arm to the body). These two possibilities are illustrated in Figures 7 and 8. When the chosen meaning indicated that the varying part of the statement (either the subject, object, or relation) was assigned to a physical object in the diagram (the ear), it was called an object mapping. When the chosen meaning indicated that the varying part of the statement was assigned to a physical relation in the diagram (the relation of the arm to the body), it was called a relation mapping.

The interesting result is that how people assigned meaning to the diagrams depended upon whether the varying part of the statement was an element (the subject or object) or a relation (the verb). When the subject or object varied, about two-thirds of participants chose meanings that were object-mappings, and when the relation varied, about two-thirds of participants chose meanings that were relation-mappings. In other words, the results of this experiment indicated that meaning is assigned to diagrams according to the similarity of relational structures. Varying elements were assigned to physical elements and varying relations were assigned to physical relations.



Fig. 6. Finally people were asked to choose one of two meanings (i.e. “Bear bites Monkey” or “Bear bites Elephant”) for each of two new diagrams. This judgment probed how people resolved the ambiguous assignment of meaning to diagram in the previous phase.

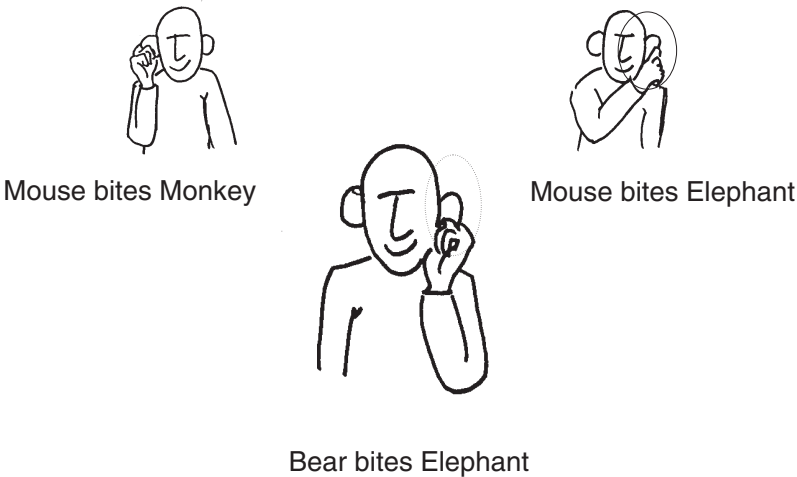


Fig. 7. The chosen meaning for a new sign was taken as an indicator of how people resolved the ambiguous assignment of meaning to the diagram. This is an example of an answer that indicates that the varying part of the statement (here the object) was assigned to a physical object in the diagram (here the left ear). This meaning assignment is an object mapping.

The results of a related experiment in which the varying relations were conjunctions (and/or) rather than active verbs indicate that this pattern of meaning assignment during diagram interpretation is true for a variety of relations, and is about something more general than just parts of speech. These results may also provide insight into how conjunctive and disjunctive relations may best be represented diagrammatically, a significant question for educators, psychologists, and computer scientists.

The sensitivity to relational structure revealed in this task does not seem to be dependent on specific experience or training, as the task is completely novel, nor on an initial object mapping in which an object is assigned to each hand. A follow-up experiment eliminated the initial step described above, in which participants were told “This hand means x,” and “This hand means y.” Instead the task began with two diagrams each paired with a meaningful statement, such as “Mouse bites Monkey”

and “Mouse bites Elephant,” or “Mouse bites Monkey” and “Mouse visits Monkey” (see Figure 5), followed by the judgment task. The results were nearly identical to the first experiment, indicating that the sensitivity to relational structure displayed in this task does not depend on any sort of priming (see Gattis, 2001d for details).

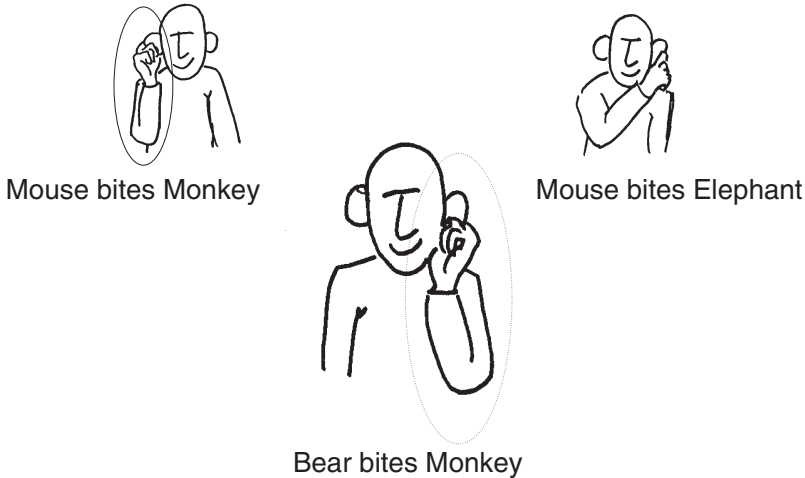


Fig. 8. This is an example of an answer that indicates that the varying part of the statement (here the object) was assigned to a physical relation in the diagram (the relation of the arm to the body, irrespective of which arm it is). This meaning assignment is a relation mapping.

3 The Role of Similarity in Representing Spatial Relations in Diagrams

While the above experiments investigated how abstract relations are represented diagrammatically, recently I have been using this task to look at how spatial relations are represented diagrammatically. Spatial relations are an interesting case because we have a great deal of experience with diagrammatic representations of spatial relations, in the form of maps, graphs, and drawings of all sorts. This rich experience set stands in contrast to our limited experience with diagrammatic representations of conjunctive and disjunctive relations or action predicates. A further reason why diagrammatic representations of spatial relations are interesting to study is because it seems likely that more than one type of similarity may be present and relevant in such diagrams, and it would be interesting to know how these forms of similarity interact. For instance, iconicity plays an important role in many maps – although the differences between maps of the same environment illustrate that iconicity is built around partial rather than complete resemblance, and iconicity is sometimes a false friend of the map-maker. Polarity also seems to influence spatial representations in which one perspective or spatial dimension is mapped onto another, as for instance when we map “front” onto “up” and “back” onto “down”, or vice versa, as we see in many maps of the world and maps of local spaces. Iconicity and polarity might seem sufficient for representing spatial relations, but maps and diagrams may also be

influenced by relational structures, for instance when a bus route is represented as a smooth line connecting points. Studies of signed languages and children's understanding of scale models also provide some evidence that relational similarity plays a role in representation of spatial relations (Emmorey, 1996; 2001; Marlzoff, DeLoache, & Kolstad, 1999). In the following two sections I will describe several recent experiments investigating whether and under what conditions relational similarity influences diagrammatic representation of spatial relations.

3.1 Representing Locative Predicates in Diagrams

The first of these experiments examined which types of similarity influence mapping of locative predicates to diagrams. First a specific meaning was assigned to each hand, as described above in Figure 4. For half of the participants, the hand-specific meanings were "car" and "office," and for half of the participants, the hand-specific meanings were "Mother" and "Father." Then, as described in Figure 5 above, two new diagrams were paired with two simple locative statements involving the object represented by the right hand. Finally just as in Figure 6 above, the judgment phase involved matching two new statements to two new diagrams. The locative statements used were "Mother is in the car" and "Mother is in the office," and "Father is in the car" and "Father is in the office." The assignment of Mother and Father to each hand was counterbalanced between subjects so that for half of the participants, the exemplars involved Mother and the probe statements involved Father, and for half of the participants it was the other way around.

While the experimental paradigm was basically the same as that described in the preceding section, it will help the reader to note that this design manipulated relational structure in a very different way. Whereas in the previous experiments relational structure was manipulated by varying which statements were given to participants in different groups (varying either the subject, the object, or the relation), in the following experiment all participants received the same set of statements, and two types of relational structure were manipulated by varying which aspect of the statement was clearly mapped in the first step of the experiment (and thus which aspect of the statement was ambiguously mapped in the following steps). This was accomplished by manipulating between subjects which aspect of the locative statement was assigned to the hands. The meanings assigned to the right and left hands were either "car" and "office," or "mother" and "father." For those participants for whom "car" and "office" were assigned to the hands, the subjects of the locative statements introduced in the second phase ("mother" and "father") were unassigned and therefore ambiguously mapped. In contrast, for those participants for whom "mother" and "father" were assigned to the hands, the locative predicates ("car" and "office") were unassigned and therefore ambiguously mapped.

The expectation was that if relational structure plays an important role in the representation of locative predicates, people would choose object mappings when the unassigned or ambiguously mapped part of the statement was subject of the sentence (Mother and Father), and relation mappings when the unassigned or ambiguously mapped part of the statement was the locative predicate ("in the car" and "in the office").

Table 1. Frequencies of each mapping pattern for diagrams paired with locative predicates, „Mother is in the car“ and „Mother is in the office“ or „Father is in the car“ and „Father is in the office.“

Expt Condition	Number of participants choosing a mapping			Binomial Test
	Object mapping	Relation mapping	N	
Subject-varying	46	26	72	p=.01
Predicate-varying	18	43	61	p=.009

As in the previous experiments, the assignment of meaning to the diagrams was determined by which aspect of the locative statement was unassigned (see Table 1). When “car” and “office” were specifically assigned to the hands, and “Mother” and “Father” were ambiguously mapped, about two-thirds of participants chose an object mapping. In contrast, when “Mother” and “Father” were specifically assigned to the hands, and “car” and “office” were ambiguously mapped, about two-thirds of participants chose a relation mapping. That actors were assigned to a spatial locus while relations were assigned to a spatial relation is confirmed by binomial tests demonstrating that for both experimental conditions, the frequencies of each mapping pattern differ significantly from an equal distribution (probabilities of the observed frequencies as determined by the binomial tests are provided in Table 1). These two patterns of response for the subject-varying condition and the predicate-varying condition were significantly different $\chi^2(1, N = 133) = 15.64, p < .001$. The overall frequency of the two mapping patterns were approximately the same: combining the two experimental conditions, participants chose object mappings and relation mappings with similar frequency.

The results of this experiment are consistent with the hypothesis that similarities of relational structure influence the interpretation of diagrams representing spatial relations. These results are compatible with the report that in signing space, objects or actors are assigned to a spatial locus, while relations between them are indicated by the use of movement (Emmorey, 1996; 2001), but point to an important difference. In this study, nouns were not always assigned to spatial loci, but rather the structural role played by a noun determined whether it was mapped to a spatial locus or a spatial relation. It appears that the nouns “car” and “office” were mapped to physical relations, not to physical objects, because they were essential parts of locative relational expressions, “in the car” and “in the office.”

3.2 Representing Spatial Prepositions in Diagrams

The results of the previous experiment demonstrated that relational structure can influence the mapping of locative statements to diagrams. The next experiments tested the generalizability of this result: three experiments investigated whether diagrams representing spatial prepositions would reveal the same sensitivity to relational structure. These experiments used the same diagrams and procedure as described in Section 2.3, with a few critical differences. Rather than action predicates or conjunctions and disjunctions, the relations used were spatial prepositions: near and far, and above and below. As described in Section 2.3, different types of relational structure were contrasted by varying either the subject, the object, or the relation.

In the first of these experiments, the spatial prepositions used were “near” and “far,” and the hands were assigned a specific meaning, as shown in Figure 4. The statements were about relations between animal characters for example, “Monkey is near to Mouse” and “Monkey is near to Elephant” or “Monkey is near to Mouse” and “Monkey is far from Mouse.” As can be seen in Table 2, when the subject or object of the statement varied, a majority of participants chose an object mapping. When the spatial preposition (“near” and “far”) varied, however, exactly half of the participants chose each possible mapping. Probabilities of the observed frequencies as determined by binomial tests are provided in Table 2.

Table 2. Frequencies of each mapping pattern for diagrams paired with statements involving spatial prepositions „near“ and „far.“

Expt Condition	Number of participants choosing a mapping			Binomial Test
	Object mapping	Relation mapping	N	
Subject-varying	31	22	53	p=.14
Object-varying	43	18	61	p=.009
Relation-varying	31	31	62	p=.55

The next experiment was identical to the previous with the difference that the initial step of assigning specific meanings to the hands was eliminated, and it was conducted in English whereas the previous experiment had been conducted in German. The results of this experiment were very similar to those of the previous experiment. Frequencies of each mapping pattern and probabilities of the observed frequencies as determined by binomial tests are provided in Table 3. One possible explanation for this pattern of results is that both iconicity and relational similarity could influence the representation, and those two forms of similarity would lead to different judgments of meaning. An iconic mapping of “near” and “far” would lead the reasoner to make an object mapping, while relational similarity of “near” and “far” would lead the reasoner to make a relation mapping. Unfortunately it is not clear how to disentangle these two influences and test this post hoc explanation of the result.

Table 3. Frequencies of each mapping pattern for diagrams paired with statements involving spatial prepositions „near“ and „far,“ and without specific assignment of meaning to the hands.

Expt Condition	Number of participants choosing a mapping			Binomial Test
	Object mapping	Relation mapping	N	
Subject-varying	21	9	30	p=.02
Object-varying	15	7	22	p=.07
Relation-varying	13	15	28	p=.42

The final experiment compared locatives involving the spatial prepositions “above” and “below.” The design was identical to the immediately previous experiment, and did not involve an assignment of specific meanings to the hands. The pattern of judgments appears to fall somewhere in-between those of the previous two experiments and those described in Section 2.3.

Table 4. Frequencies of each mapping pattern for diagrams paired with locative statements involving the spatial prepositions „above“ and „below.“

Expt Condition	Number of participants choosing a mapping			Binomial Test
	Object mapping	Relation mapping	N	
Subject-varying	15	8	23	p=.11
Object-varying	17	8	25	p=.05
Relation-varying	12	16	28	p=.29

The results of these three experiments stand in contrast to the results of previous studies. Together these results suggest that while relational similarity may influence the mapping of spatial relations to diagrams, particularly in the case of spatial relations other forms of similarity may influence diagrammatic representation as well.

4 Conclusion

A theory of diagrammatic representation requires an explicit and finely detailed account of how similarity influences the construction and interpretation of diagrams. This paper makes a small step in that direction by suggesting three types of similarity that seem to play an important role in diagrammatic representation. These three forms of similarity are iconicity, polarity, and relational similarity. While iconicity is a relatively well-known form of similarity influencing diagrammatic representation, investigations of the latter two have only recently begun. This paper has discussed experimental evidence suggesting that polarity and relational similarity do play an important role in spatial representations of abstract relations. The final section of this paper discussed how spatial relations are represented diagrammatically and presented four experiments investigating the role of similarity in representing spatial relations. The results of these four experiments are not homogenous, but as a whole they suggest that different forms of similarity may interact and even compete during diagrammatic representation. These results are interesting because they indicate that maps, graphs and drawings representing spatial relations may in some sense be representationally more complex than diagrams representing more abstract relations.

References

- Bauer, M. I., Johnson-Laird, P. N. How Diagrams can Improve Reasoning. *Psychological Science* 4 (1993) 372-378.
- Bertin, J. *Semiology of graphics* (second edition). (W. J. Berg, Trans.). The University of Wisconsin Press, Madison, WI (1983).
- Emmorey, K. The Confluence of Space and Language in Signed Languages. In P. Bloom, M. Peterson, L. Nadel, M. Garrett (eds.): *Language and Space*. The MIT Press, Cambridge (1996) 171-209.
- Emmorey, K. Space on Hand: The Exploitation of Signing Space to Illustrate Abstract Thought. In M. Gattis (ed.): *Spatial Schemas and Abstract Thought*. The MIT Press, Cambridge MA (2001a) 147-174.

- Fromkin, V., & Rodman, R. *An Introduction to Language* (sixth edition). Harcourt Brace, Fort Worth, TX (1998).
- Gattis, M. Mapping Conceptual and Spatial Schemas. In M. Gattis (ed.): *Spatial Schemas and Abstract Thought*. The MIT Press, Cambridge MA (2001a) 223-245.
- Gattis, M. Structure Mapping in Spatial Reasoning. *Cognitive Development* (in press, 2002).
- Gattis, M. Space as a Basis for Reasoning. In J. S. Gero, B. Tversky, & T. Purcell (eds.): *Visual and Spatial Reasoning in Design II*. Key Centre of Design Computing and Cognition, Sydney (2001b) 15-24.
- Gattis, M. Perceptual and Linguistic Polarity Constrain Reasoning with Spatial Representations. Manuscript in Preparation (2001c).
- Gattis, M. Mapping Relational Structure in Spatial Reasoning. Manuscript under Review (2001d).
- Gilpin, A. R., Allen, T. W. More Evidence for Psychological Correlates of Lexical Marking. *Psychological Reports* 34 (1974) 845-846.
- Hamilton, H. W., Deese, J. Does Linguistic Marking Have a Psychological Correlate? *Journal of Verbal Learning and Verbal Behavior* 10 (1971) 707-714.
- Huer, M. B. Examining Perceptions of Graphic Symbols Across Cultures: Preliminary Study of the Impact of Culture/Ethnicity. *Augmentative and Alternative Communication* 16 (2000) 180-185.
- Kotovsky, L., Gentner, D. Comparison and Categorization in the Development of Relational Similarity. *Child Development* 67 (1996) 2797-2822.
- Koul, R. K., Lloyd, L. Comparison of Graphic Symbol Learning in Individuals with Aphasia and Right Hemisphere Brain Damage. *Brain and Language* 62 (1998) 398-421.
- Markman, A. B., Gentner, D. Structural Alignment during Similarity Comparisons. *Cognitive Psychology* 25 (1993) 431-467.
- Marzolf, D. P., DeLoache, J. S., Kolstad, V. The Role of Relational Similarity in Young Children's Use of a Scale Model. *Developmental Science* 2 (1999) 296-305.
- Morford, J.P. Insights to Language from the Study of Gesture: A Review of Research on the Gestural Communication of Non-signing Deaf People. *Language & Communication* 16 (1996) 165-178.
- Peirce, C. S. *Collected Papers, Volume II: Elements of Logic* (C. Hartshorne & P. Weiss, eds.). The Belknap Press of Harvard University Press, Cambridge MA (1960/Original work published 1903).
- Stevens, S. S. *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. John Wiley, New York (1975).
- Tversky, A. Features of Similarity. *Psychological Review* 84 (1977) 327-352.
- Tversky, B. Cognitive Origins of Graphic Conventions. In F. T. Marchese (ed.). *Understanding Images*. Springer-Verlag, New York (1995) 29-53.
- Tversky, B. Spatial Schemas in Depictions. In M. Gattis (ed.): *Spatial Schemas and Abstract Thought*. The MIT Press, Cambridge (2001) 79-112.
- Vosniadou, S., Ortony, A. (eds.) *Similarity and Analogical Reasoning*. Cambridge University Press, New York (1989).
- Wertheimer, M. *Untersuchungen zur Lehre von der Gestalt, II*. *Psychologische Forschung* 4 (1923) 301-350.

Spatial Knowledge Representation for Human-Robot Interaction

Reinhard Moratz¹, Thora Tenbrink², John Bateman³, and Kerstin Fischer³

¹ University of Bremen, Center for Computer Studies
Bibliothekstr. 1, 28359 Bremen, Germany
moratz@tzi.de

² University of Hamburg, Department for Informatics
Vogt-Kölln-Str. 30, 22527 Hamburg
tenbrink@informatik.uni-hamburg.de

³ University of Bremen, FB10: Linguistics and Literary Studies
Postfach 330440, 28334 Bremen, Germany
{bateman, kerstinf}@uni-bremen.de

Abstract. Non-intuitive styles of interaction between humans and mobile robots still constitute a major barrier to the wider application and acceptance of mobile robot technology. More natural interaction can only be achieved if ways are found of bridging the gap between the forms of spatial knowledge maintained by such robots and the forms of language used by humans to communicate such knowledge. In this paper, we present the beginnings of a computational model for representing spatial knowledge that is appropriate for interaction between humans and mobile robots. Work on spatial reference in human-human communication has established a range of reference systems adopted when referring to objects; we show the extent to which these strategies transfer to the human-robot situation and touch upon the problem of differing perceptual systems. Our results were obtained within an implemented kernel system which permitted the performance of experiments with human test subjects interacting with the system. We show how the results of the experiments can be used to improve the adequacy and the coverage of the system, and highlight necessary directions for future research.

Keywords. Natural human-robot interaction, computational modeling of spatial knowledge, reference systems

1 Introduction and Motivation

Many tasks in the field of service robotics will profit from natural language interfaces that are capable of supporting more ‘natural’ styles of interaction between robot and user. Most typical scenarios include a human user instructing a robot to perform some action on some object. But a precondition for the successful performance of this kind of task is that human and robot establish joint reference to the objects concerned. This requires not only that the scene description created by the robot’s object recognition system and the visual system of the human instructor be matched, but also that the robot can successfully mediate between the two kinds of description—that is, between its internal spatial representations of the position and identity of objects and the styles

of language that humans use for achieving reference to those objects. There are two substantial problem areas facing solutions to this task: one arising out of the very non-human-like perceptual systems employed by current robots, the other out of the fact that human language users rarely employ the complete and unambiguous references to objects that might be naively expected.

In this paper, we present an experimental system that employs a computational model designed for the mapping of human and robotic systems. Based on a more detailed analysis of the results of an exploratory study which has been previously described in [Moratz et al., 2001], we show how the two problem areas at hand need to be addressed, present an expanded version of the earlier computational model that was used in the study, and open up perspectives for necessary future research.

1.1 Two Problem Areas in Achieving Spatial Reference

The first problem area, concerning the differing perceptual capabilities of humans and robots, gives rise to a range of divergences between strategies found in human-human communication and those applicable to the human-robot domain. Between humans, reference objects can usually be specified by the class name of the object. However, when the robot has no detailed *a priori* knowledge about all of the relevant objects (for example, CAD data, knowledge from a large training set), the current state of the art does not allow correct object categorization by class. Although modern automatic object recognition systems are increasingly good at identifying individual objects if the system has been trained for the specific object features, recognizing known objects is only one important aspect of successful communication between humans and robots. Very often, it is new objects in an open scenario that have to be categorized correctly in order to identify the object referred to by the speaker. This may cause severe communication problems that compromise the interaction. For example, while in human-to-human communication reference is often established by using the object category name, such as: “the key on the floor;” a corresponding natural language human-to-robot instruction that accommodates the perceptual abilities of the robot may need to be more like: “the small reflecting object on the ground, to the left of the brown box.” This is because robots have limited perceptual capabilities that often preclude accurate recognition of broadly similar objects and, moreover, may not have access to the necessary world knowledge that would identify the object by class.

The second problem area, the partiality of human strategies for achieving spatial reference, is clearly shown by work on achieving reference to objects undertaken within the field of natural language generation. Reiter and Dale (1992), for example, have shown both that the general task of producing a guaranteedly unambiguous referential expression for some object from a set of potential referents is NP-hard and that humans (perhaps as a consequence) do not in any case attempt to construct guaranteedly unambiguous references. The referential strategies adopted in natural human-human communication usually employ perceptual salience, recent mention in the discourse, and deictic modifiers (this, that, etc.) in order to achieve successful reference without requiring a solution to the problem of creating an optimal (i.e., shortest uniquely identifying) reference expression. Furthermore, since the possibility of error is in-built, interactive strategies are employed by all participants in an interaction in order to provide opportunities both

for inobtrusively exhibiting what has been understood and for smoothly correcting misunderstandings that have occurred. Only when reference is still not successful does the human interactant need to resort to explicit correction of the misunderstanding. These interactional techniques have been widely researched, particularly within the conversation analytic tradition [Schegloff et al., 1977].

The relative ‘unnaturalness’ of the second referring expression used above, which is the perceptually appropriate one for the robot, is then a direct consequence of these properties: for a natural interaction the expression sounds both over-explicit (“small reflecting ... on the ground”) and under-specific (“object” instead of “key”). Ways need to be found of ameliorating both problems if more natural interactive styles are to be achieved.

1.2 Qualitative Spatial Reference as a Communicative Device

To address these problems, a different, powerful strategy for achieving reference in human-human communication can be considered more closely. Whereas many objects may have some particular color, size or texture—which therefore give rise to more potential confusion, or ‘distractors’, for a referential expression—the position of objects is generally uniquely defining; if identified sufficiently restrictively, only one object is in a given place at a time. This could make the use of explicit positional information a good strategy for achieving unique reference in the human-robot communicative situation also. However, specifying positional information in the human-robot context also faces the above problems of mismatched perceptual systems (the robot is good at exact range-finding, humans are not) and object identification (relative positions need to reference other objects, and these objects again must be identifiable by the robot). Therefore, although more constrained, the problems above still need addressing. In this paper, we focus particularly on this use of positional information for reference to objects in human-robot interaction and attempt appropriate solutions to the accompanying problems by means of empirically deriving a spatial representation supportive of natural interaction. Qualitative spatial reference then serves as a necessary bridge between the metric knowledge required by the robot, and more ‘vague’ concepts that build the basis for natural linguistic utterances, as suggested by Hernández (1994). As the kinds of spatial representations and appropriate language forms to be adopted still need to be ascertained and evaluated empirically, we investigate this area further in an exploratory study. The results are outlined in detail in section 4.

In our scenario, a human user is asked to instruct a robot to move to one of several similar objects that are arranged in the spatial vicinity of the robot and in some cases also in the vicinity of a further, different object. The robot is equipped with a prototypical object recognition system characterized by the following features: The system can determine the metrical position (distances and angles) relative to the robot, and estimate the approximate size of an object in relation to the robot’s size—i.e., larger or smaller than the robot; it can make a coarse classification of the object’s shape (compact vs. long); and it can provide coarse colour information (ca. six to eight colour categories, although sharp distinctions between categories such as **red** and **orange** are not available). However, the system is unable to deal with gestural indications of direction. Furthermore, due to these system limitations the robot is not able to make fine distinctions between roughly similar

objects. Thus, it will not be able to identify objects correctly on the basis of a human instructor's verbal input if such input refers to fine-grained non-positional differences between the objects in question. The simple experimental configuration then forces the human user to explore other ways of referring to objects and, here, distinguishing objects on the basis of their position in space becomes a natural candidate. However, as humans in natural surroundings are not capable of providing exact metrical information about distances and angles, the objects' positions have to be referred to by qualitative information such as their relative position and other referential strategies. Ascertaining these strategies and their effectiveness was then one goal of our experimental set-up.

To formulate hypotheses about the expected user strategies in qualitative linguistic spatial reference, we can draw on previous research (e.g., [Levinson, 1996]) on human strategies for achieving such reference within naturally occurring scenarios to a certain degree. The perspective used in our scenario is, however, fundamentally different to that in most human-human interaction scenarios. In a typical experiment carried out to trigger human subjects' linguistic references, a relevant question could be: "Where is the object?". A typical answer describes the object's location by referring to its spatial relation to other available entities, such as the speaker, the hearer, or another object. In contrast, the restrictions we have seen in human-robot interaction readily create the need to refer to objects in ways which are less common in natural human-human interaction. Using the positional strategy for reference, for example, reverses this last perspective: it is not the position of an object that is unknown, but rather the *identity* of one of several entities with known positions. Thus, the issue at hand becomes: "Which of these similar objects are you referring to?". This scenario triggers strategies of linguistic reference hitherto largely ignored in the literature on spatial reference systems. We have accordingly adopted a very constrained scenario that effectively forces interaction of the kind required.

1.3 Application in a Situated and Integrated Instruction Scenario

A spatial and instructional knowledge representation serves as point of integration for the robot's language and vision mediated information. This provides for an integrative and coherent representation of objects, events and facts. Both modalities, language and vision, are then made available to the processes of understanding via a common representation level.

Central to the architecture of our experimental system is the insight that spatial instruction is:

situated: the discourse relates to a scene which can be understood using only limited previous knowledge. The visual access to a mutually perceived scene supports a state of joint attention to real-world objects.

integrated: the integration of language and vision as well as action allows a single consistent interpretation.

This architecture will be discussed below. These two central factors, situatedness and integratedness, determine the procedure used in this paper.

1.4 Related Research

Natural language is now established as a crucial component of any ‘natural’, user-friendly and appropriate interface supporting communication between computational systems and their human users. The integration of language and perception has a long tradition ([Neumann and Novak, 1983], [Wahlster et al., 1983], [Hildebrandt et al., 1995], [Moratz et al., 1995]). In the context of spatial robot-human interaction, natural language performs several particularly important interactional functions: e.g., task specification, monitoring, explanation/recovery, environment update/refinement (e.g. Stopp et al. (1994)). While mostly not focusing on spatial aspects of the interaction, many current research efforts are attempting to improve the naturalness and ease of such communication; projects such as Morpha (BMBF), SFB 360 (Bielefeld), SFB 378 (Saarbrücken) all give dialog a central place and consider it a necessary feature of robot-human interaction. Each project places different priorities and emphases on different aspects of dialog. The situation is very similar for assistance systems, such as SmartKom. Also within these projects, however, the linguistic channel is combined with interaction via graphics, gestures and the like (e.g., Streit (2001) , Lay et al. (2001) , Wahlster (2001)). This is undoubtedly important and will significantly shape the interfaces of the future. However, there are situations where the augmentation and replacement of natural language based interaction through graphical, gestural and other channels is not possible or appropriate. In such cases, as in our scenario, the only access to the robotic system the user has is the linguistic channel.

While these research areas provide valuable contributions to the questions addressed in this paper, the specific effect of a *robot* interaction partner on the linguistic and spatial choices of a human speaker has not been addressed so far. As previous studies in the related field of human-computer interaction, e.g. [Amalberti et al., 1993], [Fischer, 2000], have shown, such specific effects are highly probable, as the users’ conceptualisation of their interaction partners has considerable impact on their language. Moreover, due to the situatedness and integratedness of the communication situation the user is focused on the interaction situation itself, which increases the influence of specific situational variables. The question of which spatial reference systems are employed by speakers under which circumstances when interacting with a robot therefore still needs further exploration.

In the next section, we sketch the variability of qualitative spatial reference systems available to humans when referring to a (visually available) object’s position. Then, we describe the natural language controlled robot system that we used in our exploratory study in human-robot interaction, which is presented in section 4. The data elicited allow determining the range of spatial instructions used, showing the situatedness and integratedness of the instructions. For the experiments, a preliminary version of our computational model was used, which is described in detail in [Moratz et al., 2001]. In section 5, we present a redesign of the model, which is based on the findings of our experiment and which accounts for the range of representational choices employed by the users. Finally, we open up perspectives for future work for using human spatial reference in the interaction with robotic systems.

2 Spatial Instructions Using Intrinsic, Relative, and Absolute Reference Systems

Previous research on reference systems employed by humans for locating one object in relation to another object of a different natural kind (cf. [Levinson, 1996] and [Herrmann, 1990]) has led to the identification of three different reference systems, termed by Levinson (1996) *intrinsic*, *relative*, and *absolute*. Each of these occurs in three further variations dependent on whether the speaker, the hearer, or a third entity serves as the origin of the perspective employed. In this section, we start from this classification of spatial reference systems in order to apply it to our specific scenario involving the identification of one of several similar objects rather than the localisation of one object. Here, objects may be classified (i.e. perceptually grouped) into and referred to as groups rather than individual objects. The position of one of the objects may then be referred to by determining its position relative to the rest of the group. Such a scenario is rather typical in human-robot interaction, but has been largely ignored in previous research on linguistic spatial reference. We offer an expansion of well-established classifications of spatial reference systems to address the question how one member of a group of objects is identified¹. Furthermore, we use several applicable results from previous psycholinguistic research to formulate assumptions about which options of the variety of reference systems theoretically available to speakers can be expected to be employed by the users in our scenario.

In *intrinsic reference systems*, the relative position of one object (the *referent*) to another (the *relatum*) is described by referring to the relatum's intrinsic properties such as *front* or *back*. Thus, in a scenario where a stone (the referent) is situated in front of a house (the relatum), the stone can be unambiguously identified by referring to the house's front as the origin of the reference system: "The stone is in front of the house". In such a situation, the speaker's or hearer's position are irrelevant for the identification of the object. However, the speaker's or hearer's front or back, or, for that matter, left or right, may also serve as origins in intrinsic reference systems: "The stone is in front of you". In such cases, no further entity (such as, in our example, the house) is needed, which is why Herrmann (1990) refers to this option as *two-point localisation*.

In a scenario where *groups of objects* serve as relatum, they can only be used for an intrinsic reference system if they have an intrinsic front. For example, to identify one person in a group of people walking in one direction one could refer to "the one who walks in the front of the group."

Humans employing *relative reference systems*, or, in Herrmann's terminology, *three-point localisation*, use the position of a third entity as origin instead of referring to inbuilt features of the relatum. Thus, the stone (the referent) may be situated to the left of the house (the relatum) from the speaker's, the hearer's, or a further entity's point of view (origin): "Viewed from the hut, the stone is to the left of the house". Here, the house's

¹ Apart from the need for expansion of previous accounts, it is necessary to be very explicit about the terminology employed in our approach, as the literature on spatial reference systems is full of ill-defined, overlapping, or conflicting usages of terms. For instance, we avoid the term *deictic* as used, among others, by [Retz-Schmidt, 1988], as it has been variously used to denote contradicting concepts, see [Levinson, 1996].

front and back are irrelevant, which is why this reference system can be employed whenever the position of an object needs to be specified relative to an entity (a *relatum*) with no intrinsic directions, such as a box.

If the stone is related to a *group of* other stones, it may be situated, for instance, to the left of the rest of the group, and this may be true from the speaker's, the hearer's, or a third entity's point of view. A typical example would be, "the leftmost stone from your point of view".

In *absolute reference systems*, neither a third entity nor intrinsic features are used for reference. Instead, the earth's cardinal directions such as *north* and *south* (or, in some languages, properties such as *uphill* or *downhill* [Levinson, 1996]) serve as anchor directions. Thus, the stone may be to the north of the speaker, the hearer, or the house. Equivalently, if the stone is situated in a *group of stones*, it may be located to the north of the rest of the group. Absolute reference systems are a special case in that there is no way of labelling "origins" or "relata" in a way consistent with the other kinds of reference systems, as directions behave differently than entities.

For our experimental scenario the following initial assumptions can be made. Although humans generally use their own point of view in spatial reference, they usually adopt their interlocutor's perspective if action by the listener or different cognitive abilities on the part of the listener are involved [Herrmann and Grabowski, 1994]. Both of these factors are true in our scenario; therefore, speakers are likely to use the robot's perspective in their instructions. Furthermore, speakers will disprefer absolute reference systems as these are rarely used in natural human-human interaction in Western culture in indoor scenarios (as opposed, for instance, to Tzeltal [Levinson, 1996], [Levelt, 1996])².

Accordingly, out of the various kinds and combinations of reference systems described above, only three kinds of linguistic spatial reference are likely to be used for communication in our scenario: First, the speakers may employ an intrinsic reference system using the robot's position as both *relatum* and origin. In this case, they specify the object's position relative to the robot's front. Secondly, they can refer to a salient object, if available, as *relatum* in a relative reference system, in which case they specify the object's position relative to the salient object from the robot's point of view. Finally, they may refer to the group as *relatum* in a relative reference system. In this case, they specify the object's position relative to the rest of the group from the robot's point of view.

3 The Natural Language Controlled Robot System

The architecture of the system used for experimentation is described in detail in [Habel et al., 1999]. We summarize here the main properties of the system's components. The following components interact: the syntactic component, the semantic component, the spatial reasoning component, and the sensing and action component (see figure 1). We can see from the architecture a relatively traditional view of the role of language in robot control in that it is assumed that the human user gives sufficiently clear and unambiguous

² While, to our knowledge, this intuition has not been directly addressed experimentally, it can be derived from the literature on the kinds of spatial reference systems used by humans in diverse scenarios.

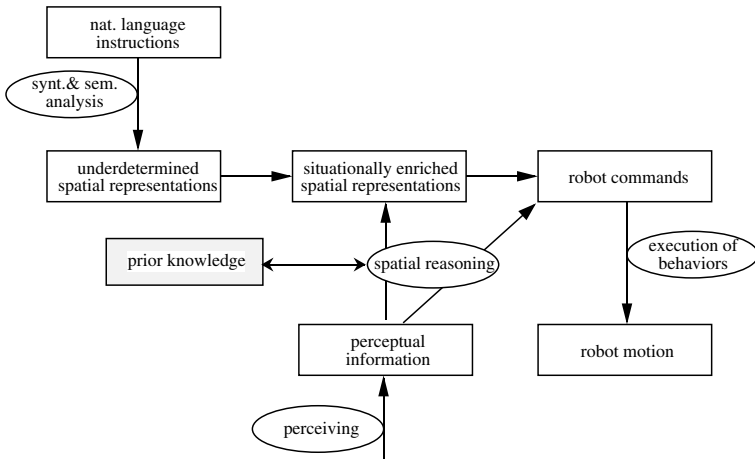


Fig. 1. Coarse architecture of a NL-instructable robot: modules and representations (from [Habel et al., 1999])

instructions for the robot to act upon; as we have suggested, for complex reference tasks this is unlikely unless the user is specifically requested to perform in this way (and even then they might not be very good at it). This simplification is appropriate for our experimental purposes, however, in that it forces the user to work through the range of referential strategies naturally available (see section 4).

The *syntactic component* is based on Combinatory Categorical Grammar (CCG), developed by Steedman and others (cf. [Steedman, 1996]). The syntactic component was developed as part of SFB 360 at the University of Bielefeld [Moratz and Hildebrandt, 1998], [Hildebrandt and Eikmeyer, 1999]. The output of the syntactic component consists of feature-value structures.

On the basis of these feature-value structures, the *semantic component* produces underspecified propositional representations of the spatial domain. In the exploratory study, this component uses a first version of our computational model of projective relations, which is described in more detail in [Moratz et al., 2001]. In section 5, we present an extended version of this model which is based on the results gained in the study. The model maps the spatial reference expressions of the given command to the relational description delivered from the sensor component.

The *spatial reasoning component* plans routes through the physical environment. To follow an instruction, the goal representation constructed by the semantic component is mapped onto the perceived spatial context.

The *sensing and action component* consists of two subcomponents: visual perception and behavior execution. The *visual perception subcomponent* uses a video camera. An important decision was to orient to cognitive adequacy in the design of the communicative behavior of the robot, using sensory equipment that resembles human sensorial capabilities [Moratz, 1997]. Therefore the camera is fixed on top of a pole with a wide angle lens looking below to the close area in front of the robot (see figure 2). The images



Fig. 2. Our Robot GIRAFFE.

are processed with region-based object recognition [Moratz, 1997]. The spatial arrangement of these regions is delivered to the spatial reasoning component as a qualitative relational description. The *behavior execution subcomponent* manages the control of the mobile robot (Pioneer 1). This subcomponent leads the robot to perform turns and straight movements as its basic motoric actions. These actions are carried out as the result of passing a control sequence to the motors.

The interaction between the components consists in a superior instruction-reaction cycle between both language components and the spatial reasoning component. Subordinate to this cycle is a perception-action cycle started by the spatial reasoning component, which assumes the planning function and which controls the sensing and action component.

An example from our application illustrates the interaction of the components and the central role of the spatial representation as follows. The command “fahre zum linken Ball” (“drive to the lefthand ball”)³ is semantically interpreted as shown in figure 3.

³ Translations are approximations and have to be treated with caution. In the mapping of spatial reference systems to linguistic expressions, there is no one-to-one correspondence between English and German.

(A) 'fahre zum linken Ball'	
(1) s:	imperativ
(2) act: type:	FAHREN
(3) agens:	GIRAFFE
(4) location: to: entity: token: ?	
(5) type:	BALL
(6) pose: relativ: xat:	LINKS

Fig. 3. Semantic interpretation

Now an object that denotes “the lefthand ball” has to be found in the perceived scene. There is a configuration of two balls one of which is to the left of the centroid of the group seen from the robot. This ball is identified as the goal of the robot. Since there is no obstacle, the action invoked will be a direct goal approach to execute the users’ command.

More complex path planning is necessary for finding paths around obstacles. To achieve this, the visual perception subcomponent has to localise the objects, and the spatial reasoning component needs to find some suitable space for movement in order to establish a qualitative route graph.

4 Exploratory Study

Our exploratory study was carried out for three primary reasons:

- Human users do not necessarily employ spatial instructions that robots can understand, and they may use strategies for spatial instruction that are different from those investigated in human-to-human communication. One aim was therefore to collect instances of spatial instructions actually employed by users in a human-robot interaction scenario.
- Since spatial instruction is situated, integrated, and involves (at least) two discourse participants, humans approach spatial instruction in an interactive way, using the situation, the actions involved as well as the kinds of sensory input available, and the possibility of interaction as a resource for their verbal instructions. We therefore aimed at working out ways in which human-robot communication is situated, integrated, and interactive.
- A third aim was to test the adequacy of the implemented version of our computational model with regard to the kinds of spatial reference systems employed by the users.

In the following section, the experimental set-up is described, section 4.2 then describes the results. Subsequent sections describe the primary uses then made of the experimental results.

4.1 Setting

The exploratory study involved a scenario in which humans were asked to instruct our Pioneer 1 robot GIRAFFE (**G**eometric **I**nfERENCE **R**obot **A**dequate **F**OR **F**loor **E**xploration, see figure 2) to move to one of several roughly similar objects. The experimenter used only pointing gestures to show the users which goal object the robot should move to; pointing was used in order to avoid verbal expressions or pictures of the scene that could impose a particular perspective, for example, a view from above. Users were instructed to use natural language sentences typed into a computer to move the robot; they were seated in front of a computer in which they typed their instructions. The users' perception of the scene was one in which a number of cubes were placed on the floor together with the robot, which was set up at a 90 degree angle or opposite to the user, as shown in figure 4. The fixed setting allows the analysis of the point of view taken by the participant depending on the instructions used. The arrangement of the cubes was varied, and in some of the settings, a cardboard box was added to the setting in order to trigger instructions referring to the box as a salient object.

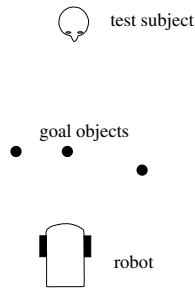


Fig. 4. The experimental setting

As outlined above, the robot can understand qualitative linguistic instructions, such as “go to the block on the right”. If a command was successful, the robot moved to the block it had identified. The only other possible response was “error”. This disabling of the natural interactive strategies of reference identification challenged users to try out many different kinds of spatial instruction to enable the robot to identify the intended aim. We were therefore able to obtain both a relatively complete indication of the kinds of strategies available to human users with respect to this task and an indication of the user's willingness to adopt them. 15 different participants carried out an average of 30 attempts to move the robot within about 30 minutes time each. Altogether 476 instructions were elicited.

4.2 Experimental Results

Throughout the experiments, the participants employed the robot's perspective, i.e., there were virtually no instructions in which the user expected the robot to use a reference system based on the speaker or a further object as origin (except for one case in which

after a mistake the user explicitly stated that she assumed the robot to be using her point of view). Furthermore, whenever the users referred to the goal object, they overwhelmingly used basic level object names such as *Würfel* [**cube**], and there was also a very consistent usage of imperatives rather than other, more polite, verb forms.

However, the participants in the experiment nevertheless showed considerable variation with regard to the instructional strategies employed. Half of the participants started by referring directly to the goal object, using instructions such as *fahr bis zum rechten Würfel* [**drive up to the right cube**]. When instructions of this kind were not successful—because of orthographic, lexical, or syntactic problems—the participants turned to directional instructions; if successful, they re-used this goal-naming strategy in later instructions. The other half of the participants started by describing the direction the robot had to take, for instance, *fahr 1 Meter geradeaus* [**drive 1 meter straight ahead**]. If they were unsuccessful with this type of instruction, some users turned to decomposing the action into even more detailed levels of granularity, using instructions such as *Dreh dein rechtes Rad* [**turn your right wheel**].

This pattern of usage reveals an implicational hierarchy among the adopted strategies. On reaching a failure, users would change their strategy only in the direction of expected ‘simplicity’; they would not attempt a strategy with expected ‘higher’ complexity. Thus, a fixed order of instructional strategies became apparent which can be roughly characterized as Goal - Direction - Minor actions. This is an important result for designing human-robot interaction—not least because the notion of ‘simplicity’ maintained by a user need not relate at all to what is actually simpler for a robot to comprehend and carry out. Thus attempts on the part of the user to provide ‘simpler’ instructions may in fact turn out to confuse rather than aid the situation.⁴ Such mismatches can therefore lead to insoluble dialogic problems that are particularly frustrating for users, since they believe (mistakenly) that they are making things easier for the robot. Thus, in the future dialogue components will need to be designed that can detect such a situation and then correct the user’s underlying assumptions unobtrusively.

In the following, we analyse in detail the kinds of spatial reference systems employed in these different kinds of instruction. As our aim was to explore the range of instructions employed by the users, and to analyse their instructional strategies on a qualitative level, we did not attempt to work out user preferences quantitatively, using statistical measures. However, for illustration of the tendencies we worked out, we add the absolute numbers of occurrence.

Goal Instructions (183 Occurrences). Spatial instructions indicating the position of the goal object are identified as bounded linear oriented structures in [Eschenbach et al., 2000]. They include directional prepositional phrases specifying the end of the path. Out of the 183 linguistic instructions collected in our experiment that refer directly to the goal object, 102 utterances use the group as a whole as relatum, identifying the intended object by its position relative to the other objects in the group. 69 of these 102 group-based references used a particular expression schema consisting of an imperative

⁴ A related instance of this problem has also been noted when attempting to have users produce more intelligible speech. This can easily lead a user to ‘hyper-articulate’ which reduces the reliability of speech recognition still further [Oviatt et al., 1998].

combined with a locative directional adjunct specifying relative position; as, for example, in *Fahr zum linken Würfel* [**Drive to the lefthand cube**] where the locative adjunct gives the relative position of the cube in the group to which it belongs. The lexical slots for the verb and object in this schema were varied, as were the positional adjective of the locative adjunct—yielding “mittleren” [**middle**], “hinteren” [**back**], “vorderen” [**front**] in addition to “linken” [**left**].

For some situations, besides the cubes used as goal objects the setting included a further object, namely a cardboard box which could be used as a reference object. In 19 cases of the 43 instructions uttered in situations where this salient object was present, the cardboard box was used for a relative reference system with the salient object as relatum. Here, the syntactic structure used most often is also quite stable: an imperative and two hypotactic adjuncts are used, with the subordinated adjunct identifying the relatum’s position relative to the adjunct specifying the reference object, as in: *geh zum Würfel rechts des Kartons* [**go to the cube to the right of the box**].

The robot’s intrinsic properties are used for instruction in altogether 42 of the 183 goal-oriented instructions, using various linguistic expressions such as *Fahr zum Würfel rechts von dir* [**Drive to the cube to your right**]. Although the orientation of the robot is not stated explicitly in these commands, the speakers could not use an expression like “to your right” without assuming a front of the robot.

Altogether, these results correspond to the expectations we outlined in section 2. Those users who referred to the goal object all employed the three kinds of reference systems expected, and they consistently used the robot’s perspective (which is actually a more homogeneous usage than we might expect). Strikingly, in all of the goal instructions except for those employing the robot’s intrinsic properties, the users failed to specify the point of view they employed, rendering the instructions formally ambiguous with regard to the variability of origins but, we would claim, appropriate within the particular situated interaction.

Direction Instructions (210 Occurrences). In altogether 210 instructions, the goal object is not specified directly, but a direction of movement is indicated. In more than half of these instructions, a verb of locomotion such as *fahre* [**drive**] or *rolle* [**roll**] is used; the others simply specify the direction itself. This variability does not reflect any relevant semantic differences (the only way the robot can move at all is by using its wheels) and are therefore not discussed further here. Other verbs of motion such as verbs of transport (“bring”), change of position (“enter”) and caused change of position (“put”) (cf. [Eschenbach et al., 2000]) do not occur in this simple scenario. Directional instructions indicate unbounded linear oriented structures, as only the initial step of an intended goal-directed path is expressed. No further steps occur in this scenario for lack of reaction by the robot. As an exception, two instructions may be combined in one utterance (see below), but these still do not include a goal, i.e., the structures are still unbounded. In more than half of the directional instructions (141 out of 210), the intrinsic point of view of the robot is used as origin of a reference system which employs the principal directions as defined in [Eschenbach, 2001]. In 78 of these cases, these principal directions are employed without modifications, as in: *vorwärts* [**forward**], and *gehe nach links* [**go to the left**]. “Vorwärts” expresses the standard orientation of a

body during motion, i.e., the alignment of the object order of the path with the intrinsic front-back axis of the robot (cf. [Eschenbach, 2001]). Several users employed the earth's cardinal directions (12 occurrences) rather than relying on the principal directions based on the robot's physical properties, as in *Gehe nach Norden* [**Go to the North**]. Altogether, in almost half (90 out of 210) of the directional, non-goal specifying instructions, the users indicated an unmodified principal or absolute direction to make the robot move, obviously leaving further specifications of the path for later instructions.

Nevertheless, many users seemed to assume that the intended goal was not directly accessible by simply moving in one of these cardinal directions. Thus, in 32 instructions the angle in which the robot should move is specified more exactly, using either quantitative (8 occurrences) measures such as *20 Grad nach rechts* [**20 degrees to the right**] or qualitative (24 occurrences) specifications, for instance, *geradeaus etwas rechts fahren* [**drive forward somewhat to the right**]. One-third of these instructions employed a combination of either a principal direction and an angle (in quantitative usages), or two principal directions (in qualitative usages). Some users explicitly divided such a combination into two partial instructions (4 occurrences) which were to be carried out one after the other, as in *gehe vorwärts dann nach rechts* [**move forward then to the right**].

Some users indicated the length of the intended path, using either quantitative (18 occurrences) measures such as *Fahre 1 meter geradeaus* [**Drive forward 1 meter**], or qualitative (8 occurrences) expressions such as *Fahre ein wenig nach vorn* [**Drive a bit forward**]. Interestingly, in contrast to the findings on angle specifications, in this case the quantitative instructions outweighed the qualitative ones. One user tried out an instruction specifying not only the direction but also the length of time during which the robot was supposed to move in that direction: *Fahre 1 Sekunde vorwärts* [**Drive forward 1 second**]. Some of the instructions (52 occurrences) relied on a different, salient entity (a landmark) available in the room for specifying the intended path rather than relying on the principal directions determined by the robot's intrinsic properties. Of these 52 instructions, 46 referred to the cardboard box which was available only in some of the scenarios, as in: *umfahre den Kasten* [**drive around the box**]. Mostly, these instructions (in contrast to the goal-based instructions) do not command the robot to move *to* the box, but rather around it, behind it, or beside it. Thus, it is linguistically expressed that the box is not itself the intended goal. The other 6 instructions used entities located at a greater distance from the robot to specify the intended direction, as in *Fahre zur Wand* [**Drive to the wall**].

Finally, in a few (4) instructions the users left it to the robot to decide about the correct orientation, as in *Fahre im Kreis* [**Drive in a circle**].

Minor Action Instructions (83 Occurrences). The remaining 83 instructions did not specify either the goal object or a direction in which the robot should move, but instead decomposed the action into minor activities. In 28 of these instructions, the users did not command the robot to **move** in a direction, but rather to change its orientation into a specific direction, as in *dreh dich nach rechts* [**turn to the right**].⁵ About half of these instructions involved qualitative, the other half quantitative measures. 29 instructions

⁵ These are not counted as directional movement instructions as they express an action on a finer level of granularity, leaving out locomotion.

indicated that the robot should move, but were confined to the verbs of locomotion, such as *Fahren* [**Drive**]. The remaining 26 instructions reflected the users' individual, sometimes rather desperate attempts to communicate with the robot at all, as exemplified by utterances such as *Tu was* [**Do something**] and *Schalte den Motor ein* [**Turn on the engine**].

5 A Revised Computational Model for the Spatial Human-Robot Interaction Scenario

The experiments described above, which were carried out using a previous version of the system, provided several valuable clues as to how a new system could be designed. A new system is currently being set up, and instead of using the GIRAFFE platform, this new system will use the Sony AIBO robotic system (see figure 5). Because of its animal-like shape, the AIBO might be perceived as a more natural communication partner. In addition, the AIBO is also well suited for gathering data on robots used in an entertainment context.



Fig. 5. Legged AIBO robot.

However, the four-legged robot AIBO and the GIRAFFE differ regarding their respective perceptual abilities (field vs. survey perspective; orientation knowledge vs. position - i.e., orientation and distance - knowledge), which might trigger various kinds of interesting communication problems. Because the AIBO camera in its head is closer to the ground than that of GIRAFFE, the AIBO is unable to calculate precise distances to unrecognized objects. Thus AIBO has only orientation knowledge available to it, and this knowledge has to tally with the survey knowledge provided by the human instructor.

By changing its position, the AIBO acquires new perspectives and further orientation information on the scene. A spatial inference engine combines the information from the AIBO's different viewpoints, along with the survey knowledge provided verbally by the human instructor, to build up a depiction of the environment. In order to draw spatial inferences using the spatial inference engine, the verbal description provided by the human instructor must first be transferred into a spatial reasoning calculus (QSR calculus). For our system, we employ the TPCC calculus, introduced in the current volume (see Moratz, Nebel, Freksa (2002)).

Given the results of our experiments, and building on the general results from psychology and psycholinguistics on spatial expressions in human-to-human communication that we summarized in section 2 above, it was possible to design a level of representation that provides our robot with a model of the verbal strategies of spatial instructions produced by users in the experimental scenario. This model consists of two parts: first, a knowledge base representing the coarse structure and links to general world knowledge (section 5.1); second, a representation capturing the fine grained positional information (section 5.2) represented using the TPCC calculus [Moratz et al., 2002]. The knowledge base offers a blueprint from which individual spatial instructions can be derived as particular instances. Such instances then provide the necessary link between the language input module and the navigation module presented in section 3 above.

5.1 The Semantic Structure of Spatial Instructions for Mobile Robots

The representation formalism we adopt is derived from the ERNEST system ([Niemann et al., 1990], [Kummert et al., 1993], [Moratz, 1997]). ERNEST is a semantic network formalism in the KL-ONE tradition, providing a subset of representation and inference capabilities relevant for robotic reasoning. It can be used for the representation of concepts and the relationships between them and has already been applied successfully in the context of integration of linguistic and perceptive knowledge [Hildebrandt et al., 1995]. Since we do not use the inference mechanisms but only the declarative component we can work with a simplified version of ERNEST, which we present here in a short sketch.

The primary elements of an ERNEST semantic network are concepts, their attributes and the relations between concepts. These are usually represented as nodes, their internal structures and links between nodes respectively. We use two types of nodes:

- A *concept* can represent a class of objects, events or abstract conceptions.
- An *instance* is understood as the concrete realisation of a concept in the input data; i.e. an instance is the copy of a concept by which the general description is replaced by concrete values.

Subordinate features of a concept, such as the size of an object or its colour, are represented by means of attributes. Concepts are therefore entities with internal structure. Features of concepts that are important for the domain are represented as links to other concepts. ERNEST supports the following standard link types:

- Through the link type *role*, two concepts are connected with each other if one concept is understood as a prerequisite of the other.

- Through the link type *specialisation* and a related inheritance mechanism, a special concept is stated to inherit all attributes and roles of the general one.

The knowledge present in the semantic network is utilized by creating instances. This process requires that a complex object be recognized as an instance of a concept, which in turn requires that all its necessary roles can be recognized.

The experimental results indicate that certain kinds of information concerning spatial directions commonly occur together and others less so. This was modeled as the semantic network fragment shown in figure 6. In the figure, specialisation links are oriented horizontally and role links are oriented vertically. Optional role links (i.e., the cardinality range includes zero) are shown dashed in the figure. The three main types of instructions found empirically constitute the three specializations shown for the concept ‘drive instruction’; the presence of a goal-object (as a subconcept of spatial-object) and of a landmark (a further subconcept of spatial-object) in their respective instruction types are shown by the vertical role links. The obligatory relationship expressed between ‘relative position’ and ‘orientation’ is the link to the ‘projective-expression’ concept which is the interface to the model of projective relations presented in the next subsection.

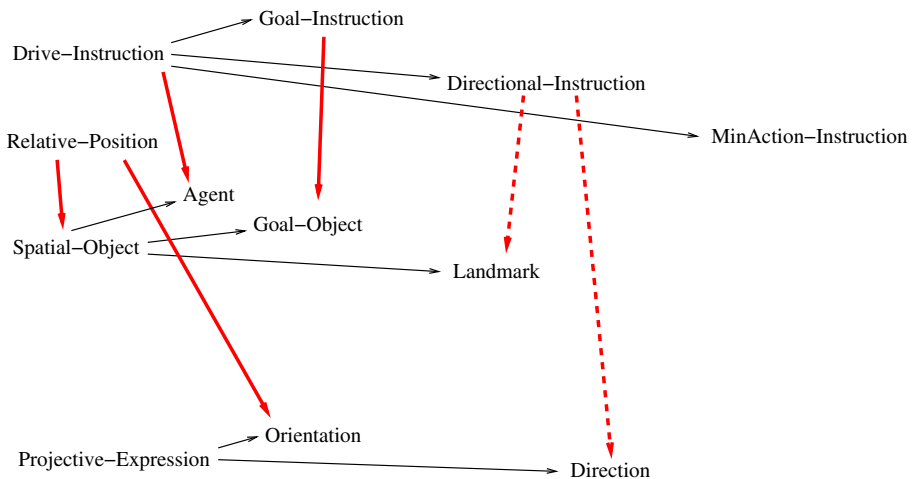


Fig. 6. Knowledge base for the semantic structure of spatial instructions.

Instances formed from these concepts interface directly with the robot’s control components. Thus, the recognition of a linguistic instruction is responded to by a corresponding action on the part of the robot. Particular instances also have information added via the robot’s perceptive apparatus; for example, exact position relative to the robot and basic attributes of colour and size as mentioned above. We will return to some further possible uses of this additional information in section 6 below.

5.2 The Interpretation of Projective Relations

An essential aspect of the robot's ability to execute instructions is its interpretation of the spatial relations specified between objects functioning as landmarks or relatum and the goal objects. The experimental results have a number of consequences for our model of the projective relations and their uses. The computational model shown in figure 6 represents the different kinds of reference systems required for interpreting linguistic references according to the three options outlined in section 2 and for handling the corresponding instructions. Note that our empirical results already allow us to exclude several theoretically possible alternatives that were not, in fact, selected as strategies by our experimental participants: for example, intrinsic and relative reference systems employing either the speaker or a salient object as origin.

The projective expressions are then further resolved as follows. To model reference systems that take the robot's point of view as origin, all objects are represented in an arrangement resembling a plan view (a scene from above). This amounts to a projection of the objects onto the plane $\mathcal{D}$ on which the robot can move. The projection of an object O onto the plane $\mathcal{D}$ is called $p_{\mathcal{D}}(O)$. The center μ of the projected area can be used as point-like representation O' of the object O : $O' = \mu(p_{\mathcal{D}}(O))$. The reference axis is then a directed line through the center of the object used as relatum (see figure 7), which may be the robot itself, the group of objects, or other salient objects.

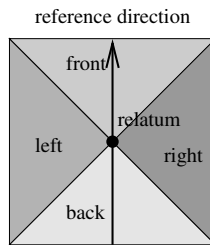


Fig. 7. Relatum and reference direction

The partitioning into sectors of equal size is a sensible model for the directions “links” (left), “rechts” (right), “vor” (front) and hinter (back) relative to the relatum. However, this representation only applies if the robot serves as both relatum and origin. If a salient object or the group is employed as the relatum, front and back are exchanged, relative to the reference direction [Herrmann, 1990]. The result is a qualitative distinction, as suggested, for instance, by Hernandez (1994). An example for this configuration is shown in figure 8. In this variant of relative localisation, the “in front of” sector is directed towards the robot.

In cases with a group of similar objects, the centroid of the group serves as virtual relatum. Here the reference direction is given by the directed straight line from the robot

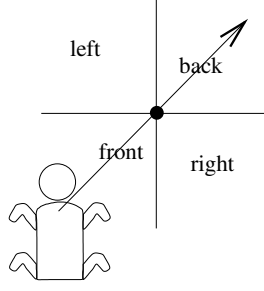


Fig. 8. Relative reference model

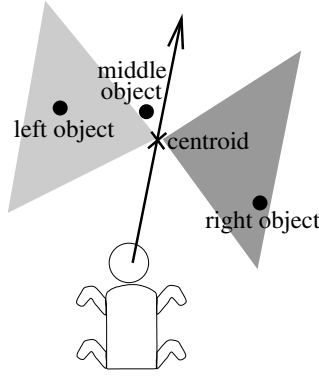


Fig. 9. Group based references

center to the group centroid. The object closest to the group centroid can be referred to as the “middle object” (see figure 9).

For combined expressions like “links vor” (left in front of) vs. precise expressions like “genau vor” (straight in front of) we use the partition presented in figure 10. This partitioning can account for the projective expressions used for the orientation in goal instructions as well as the directions in directional instructions (see figure 6 above), in which the robot’s position and physical orientation provide the basis for determining the intended reference direction.

To define the partitions formally, we refer to the angle ϕ between the reference direction and the straight line from the relatum to the referent, or, respectively, the denoted direction.

$$\begin{aligned}
 \text{referent vor relatum} &:= -\pi/4 \leq \phi \leq \pi/4 \\
 \text{referent links relatum} &:= \pi/4 < \phi < 3/4\pi \\
 \text{referent hinter relatum} &:= 3/4\pi \leq \phi \leq 5/4\pi
 \end{aligned}$$

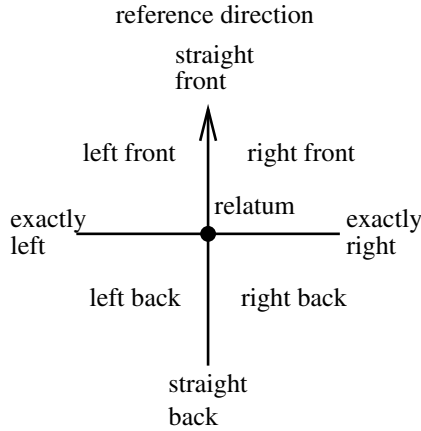


Fig. 10. Model for combined expressions

<i>referent rechts relatum</i>	$:= -\pi/4 > \phi > -3/4\pi$
<i>referent links – vor relatum</i>	$:= 0 < \phi < \pi/2$
<i>referent links – hinter relatum</i>	$:= \pi/2 < \phi < \pi$
<i>referent rechts – vor relatum</i>	$:= 0 > \phi > -\pi/2$
<i>referent rechts – hinter relatum</i>	$:= -\pi/2 > \phi > -\pi$
<i>referent genau – vor relatum</i>	$:= \phi = 0$
<i>referent exakt – links relatum</i>	$:= \phi = \pi/2$
<i>referent genau – hinter relatum</i>	$:= \phi = \pi$
<i>referent exakt – rechts relatum</i>	$:= \phi = -\pi/2$

The partitions described above exactly correspond to the acceptance areas used in the QSR calculus TPCC (this volume [Moratz et al., 2002]). With the aid of these acceptance areas, the instructor’s verbal spatial description information can be matched to the perceptually captured local view information from the AIBO. One difficulty inherent in this process is that the local view information captured by the AIBO contains only orientation knowledge, lacking distance information. However, the knowledge represented in TPCC can be combined using constraint propagation, and thus it is possible to generate survey knowledge from local knowledge.

6 First Steps Towards More Natural Interaction

The design and implementation of the mobile robot GIRAFFE reported so far has already achieved the integration of several different informational modalities. Linguistic input, perception and robot action all combine in the robot’s interpretation and execution of the instructions it receives. The implemented model performs adequately in that its primary

behavioral mode, following goal-centered instructions, corresponds to the instruction strategy most preferred by users. Users overwhelmingly employed the robot's perspective and most of the spatial reference systems employed corresponded directly to those implemented so that successful communication was achieved. Moreover, based on the fact that there were situations in which other strategies were employed, such as directional instructions or specifications of minor actions, our inclusion of these in the model will allow successful interaction here also.

The experimental results of course raise many more issues. In particular, we consider in this section a more sophisticated use of spatial representation in order to allow successful operation in more demanding circumstances. The correct achievement of reference in human-human interaction is often more negotiated and interactively mediated than was supported by our experimental scenario. However, given the representation uncovered, we are now exploring ways of using interaction to clarify underlying misconceptions on the part of the user such as that which led some of our test persons to believe that they could not directly refer to the intended goal object; and to allow more robust and powerful recognition by the robot of a user's instructions. This can be clarified with some simple examples. Goal objects can currently be recognized on the basis of the linguistic input to the system only when there are not too many competing potential referents. If, for instance, there are several cubes 'to the left' within a group of cubes, then simple reference may fail. Moreover, the very fact that there is a more complex situation for which a user must construct an appropriate referring expression can lead to the production of language that falls outside the limited expressive power of the semantic/pragmatic interpreter or even to expressions that are not strictly correct as referring expressions.

We can improve on this situation by making sure that the user's referring acts are embedded in a discourse—in particular, in an ongoing interaction initiated by the robot and which defines the rules of the game. Thus, if the robot first informs the user what it can perceive in a scene, then the terms and perspectives available for reference are already constrained favorably for the robot's subsequent interpretation. This requires that the robot be in a position to *verbalize* its scene perception. The kind of domain knowledge representation for our newly designed AIBO scenario sketched in figure 6 already goes a considerable way towards this. Standard techniques from the area of natural language generation (e.g., [Horacek, 2001]) work on collections of instances organized in terms of domain conceptual hierarchies such as the one given here in order to produce natural language descriptions of the requested content. Part of this work involves aggregating the objects present into referring expressions that allow the user to identify what is being described (cf. [Bateman, 1999]). These referring expressions can then already be used to suggest to users particular ways of describing the goal objects of their required instructions.

In a simple case, for example, there may be a scene in which there are several similar cubes within the same spatial sector, but where those cubes differ with respect to some other attribute: two may be red, another blue. Standard aggregation techniques can pick out the differing attributes and use these to determine appropriate referring expressions: thus, we can ascertain that 'the blue cube' is sufficient to identify the one blue cube in question, while 'the red cube' will need further elaboration (e.g., by the projective

relations described above). As the situation to be described becomes more complex, correspondingly more complex referring expressions may be produced that are limited in the practical ways already investigated in detail in work such as that of Reiter and Dale (1992). In particular, aggregation can establish particular groups in the discourse to serve as the relatum in projective expressions of the kind illustrated above. A robot-produced utterance such as “To my left there are two red cubes and one blue cube” introduces in addition to the three objects with the sector ‘left’ a subgroup consisting of just the two red cubes. Subsequent reference can use this just as the perceptually defined groups were used above: e.g., “the rightmost red cube”. Within an interaction, interpretation of this expression can be constrained to the group of red cubes introduced by the robot at that point, rather than referring to the entire set of red cubes possibly available in the scene at large. Reference thus becomes both interactional and situated as is natural in human-human interaction.

In a different scenario, human users may be expected to use spoken language to address the robot, rather than to type their instructions. While such a scenario at first sight is undoubtedly more natural to the users, it raises a range of different problems, such as those occurring when the speech input is not recognised correctly by the system, or those caused by user expectations about the system’s facilities. Compared to our scenario, it is by no means clear which kinds of language users would employ if required to talk rather than to type. It is well-known that spoken language differs in many respects from written language; also depending on other situational factors [Biber, 1988]. Further research is needed to explore this and other kinds of variation with regard to the enhanced human-robot interaction scenario with our new robot AIBO which we are now implementing.

7 Conclusion

In this paper, we have described an implemented mobile robot system that follows simple instructions given by its human users. We have investigated empirically the kinds of instructions that users employ and have provided a computational model of these strategies as a level of spatial instruction knowledge representation that interfaces between the linguistic input provided to the robot and the robot’s sensing and action component. This implemented version of the system was demonstrated to perform in an adequate way, but only in a relatively simple set of possible task scenarios. We then briefly sketched a current direction of research in which we are building on the explicit spatial instruction model in order to provide more interactive linguistic behavior. This will feed into a further round of empirical investigation, which will evaluate the effectiveness of the functionalities provided. We have suggested that this is a necessary and beneficial step towards achieving more robust and natural interactional styles between humans and mobile robots.

Acknowledgement

The authors would like to thank Carola Eschenbach, Christian Freksa, Christopher Habel and Tilman Vierhuff for interesting and helpful discussions related to the topic of the paper. We thank Bernd Hildebrandt for constructing the parser. And we would like

to thank Jan Oliver Wallgrün, Stefan Dehm, Diedrich Wolter and Jesco von Voss for programming the robot and for supporting the experiments. Also many thanks to Christie Manning for helpful comments on our paper.

References

- Amalberti et al., 1993. Amalberti, R., Carbonell, N., and Falzon, P. (1993). User Representations of Computer Systems in Human-Computer Speech Interaction. *International Journal of Man-Machine Studies*, 38:547–566.
- Bateman, 1999. Bateman, J. A. (1999). Using aggregation for selecting content when generating referring expressions. In *Proceedings of the 37th. Annual Meeting of the Association for Computational Linguistics (ACL'99)*, pages 127–134, University of Maryland. Association for Computational Linguistics.
- Biber, 1988. Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press, Cambridge.
- Eschenbach, 2001. Eschenbach, C. (2001). Contextual, Functional, and Geometric Features and Projective Terms. In *Proceedings of the 2nd annual language & space workshop: Defining Functional and Spatial Features*, University of Notre Dame.
- Eschenbach et al., 2000. Eschenbach, C., Tschander, T., Habel, C., and Kulik, L. (2000). Lexical Specification of Paths. In Freksa, C., Habel, C., and Wender, K. F., editors, *Spatial Cognition II*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin.
- Fischer, 2000. Fischer, K. (2000). What is a situation? In *Proceedings of G talog 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue*, pages 85–92.
- Habel et al., 1999. Habel, C., Hildebrandt, B., and Moratz, R. (1999). Interactive robot navigation based on qualitative spatial representations. In Wachsmuth, I. and Jung, B., editors, *Proceedings Kogwis99*, pages 219–225, St. Augustin. infix.
- Hernández, 1994. Hernández, D. (1994). *Qualitative representation of spatial knowledge*. Lecture Notes in Artificial Intelligence. Springer Verlag, Berlin, Heidelberg, New York.
- Herrmann, 1990. Herrmann, T. (1990). Vor, hinter, rechts und links: das 6h-modell. psychologische studien zum sprachlichen lokalisieren. *Zeitschrift für Literaturwissenschaft und Linguistik*, 78:117–140.
- Herrmann and Grabowski, 1994. Herrmann, T. and Grabowski, J. (1994). *Sprechen: Psychologie der Sprachproduktion*. Spektrum Verlag, Heidelberg.
- Hildebrandt and Eikmeyer, 1999. Hildebrandt, B. and Eikmeyer, H.-J. (1999). *Sprachverarbeitung mit Combinatory Categorical Grammar: Inkrementalität & Effizienz*. SFB 360: Situierete Künstliche Kommunikatoren, Report 99/05, Bielefeld.
- Hildebrandt et al., 1995. Hildebrandt, B., Moratz, R., Rickheit, G., and Sagerer, G. (1995). Integration von bild- und sprachverstehen in einer kognitiven architektur. In *Kognitionswissenschaft*, volume 4, pages 118–128, Berlin. Springer-Verlag.
- Horacek, 2001. Horacek, H. (2001). Textgenerierung. In Carstensen, K.-U., Ebert, C., Endriss, C., Jekat, S., Klabunde, R., and Langer, H., editors, *Computerlinguistik und Sprachtechnologie – Eine Einführung*, pages 331–360. Spektrum Akademischer Verlag, Heidelberg.
- Kummert et al., 1993. Kummert, F., Niemann, H., Prechtel, R., and Sagerer, G. (1993). Control and explanation in a signal understanding environment. *Signal Processing, special issue on 'Intelligent Systems for Signal and Image Understanding'*, 32:111–145.
- Lay et al., 2001. Lay, K., Prassler, E., Dillmann, R., Grunwald, G., H gele, M., Lawitzky, G., Stopp, A., and von Seelen, W. (2001). MORPHA: Communication and Interaction with Intelligent, Anthropomorphic Robot Assistants. In *International Status Conference: Lead Projects Human-Computer-Interaction*, Saarbruecken, Germany.

- Levelt, 1996. Levelt, W. J. M. (1996). Perspective Taking and Ellipsis in Spatial Descriptions. In Bloom, P., Peterson, M., Nadel, L., and Garrett, M., editors, *Language and Space*, pages 77–109. MIT Press, Cambridge, MA.
- Levinson, 1996. Levinson, S. C. (1996). Frames of Reference and Molyneux's Question: Crosslinguistic Evidence. In Bloom, P., Peterson, M., Nadel, L., and Garrett, M., editors, *Language and Space*, pages 109–169. MIT Press, Cambridge, MA.
- Moratz, 1997. Moratz, R. (1997). *Visuelle Objekterkennung als kognitive Simulation*. Disk 174. Infix, Sankt Augustin.
- Moratz et al., 1995. Moratz, R., Eikmeyer, H., Hildebrandt, B., Kummert, F., Rickheit, G., and Sagerer, G. (1995). Integrating speech and selective visual perception using a semantic network. *Proc. AAAI-95 Fall Symposium on Computational Models for Integrating Language and Vision*, pages 44–49.
- Moratz et al., 2001. Moratz, R., Fischer, K., and Tenbrink, T. (2001). Cognitive Modeling of Spatial Reference for Human-Robot Interaction. *International Journal on Artificial Intelligence Tools*, 10(4):589–611.
- Moratz and Hildebrandt, 1998. Moratz, R. and Hildebrandt, B. (1998). *Deriving Spatial Goals from Verbal Instructions - A Speech Interface for Robot Navigation -*. SFB 360: Situierete Künstliche Kommunikatoren, Report 98/11, Bielefeld.
- Moratz et al., 2002. Moratz, R., Nebel, B., and Freksa, C. (2002). Qualitative spatial reasoning about relative position: The tradeoff between strong formal properties and successful reasoning about route graphs. *this volume*.
- Neumann and Novak, 1983. Neumann, B. and Novak, H.-J. (1983). Event models for recognition and natural language description of events in real-world image sequences. In *IJCAI 1983*, pages 643–646.
- Niemann et al., 1990. Niemann, H., Sagerer, G., Schröder, S., and Kummert, F. (1990). ERNEST: a semantic network system for pattern understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):883–905.
- Oviatt et al., 1998. Oviatt, S., MacEachern, M., and Levow, G.-A. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication*, 24:87–110.
- Reiter and Dale, 1992. Reiter, E. and Dale, R. (1992). A fast algorithm for the generation of referring expressions. In *Proceedings of the fifteenth International Conference on Computational Linguistics (COLING-92)*, volume I, pages 232–238, Nantes, France. International Committee on Computational Linguistics.
- Retz-Schmidt, 1988. Retz-Schmidt, G. (1988). Various Views on Spatial Prepositions. *AI Magazine*, 9(2):95–105.
- Schegloff et al., 1977. Schegloff, E., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organisation of repair in conversation. *Language*, 53:361–383.
- Steedman, 1996. Steedman, M. (1996). *Surface Structure and Interpretation*. MIT Press, Cambridge, MA.
- Stopp et al., 1994. Stopp, E., Gapp, K.-P., Herzog, G., Laengle, T., and Lueth, T. C. (1994). Utilizing Spatial Relations for Natural Language Access to an Autonomous Mobile Agent. *Künstliche Intelligenz*, pages 39–50.
- Streit, 2001. Streit, M. (2001). Why Are Multimodal Systems so Difficult to Build? - About the Difference between Deictic Gestures and Direct Manipulation. In Bunt, H. and Beun, R.-J., editors, *Cooperative Multimodal Communication*. Springer-Verlag, Berlin, Heidelberg.
- Wahlster, 2001. Wahlster, W. (2001). SmartKom: Towards Multimodal Dialogues with Anthropomorphic Interface Agents. In *International Status Conference: Lead Projects Human-Computer-Interaction*, Saarbrücken, Germany.
- Wahlster et al., 1983. Wahlster, W., Marburger, H., Jameson, A., and Busemann, S. (1983). Over-answering yes-no questions: Extended responses in a nl interface to a vision system. In *IJCAI 1983*, pages 643–646.

How Many Reference Frames?

Eric Pederson

University of Oregon, Linguistics Department
epederso@darkwing.uoregon.edu

Abstract. There is considerable cross-disciplinary confusion concerning the taxonomy of reference frames and no standard of comparison for reference frame usage exists to allow reliable comparison of cross-linguistic, cross-cultural, and task-specific variation. This paper proposes that we examine reference frame selection in terms of the underlying component operations. The selection of these operations can be mapped out in a multi-dimensional space defined in terms of the scalar properties of the reference objects and their relationship to the speaker/viewer.

1 Introduction

The notion of spatial reference frames (RFs) has spread out from psychology into linguistics, computer science, and related fields. Each domain, however, finds itself tackling the problem of identifying RFs from different perspectives and drawing on different sets of eligible data. This has the invigorating effect of cross-disciplinary fertilization of ideas, but at the cost of some clarity concerning the notion of “reference frame” itself, as each discipline uses the term in its own ways. This paper queries whether we might hope for some pan-disciplinary constancy in RF descriptions and along the way asks the perhaps simpler question of how many reference frames are actually in use and when they are available.

For a fairly standard working definition, an RF is taken to be the imposition of some measure of *orientation* such that an entity’s location can be indicated with respect to some landmark object and/or observer. This will exclude simple mention of proximity or contact with a landmark when that description fails to specify any angular relationship between the entity and its landmark.

Everyone seems to agree that multiple RFs are cognitively and linguistically available, but which ones are available to whom remains controversial. How many and which RFs are necessary seems to depend heavily on not only the context of experience and the nature of the task being evaluated, but also on the scientific discipline of the analyst. While this variation is non-arbitrary, it paints a confused picture of human spatial operations. Further, the selection of RFs appears to be profoundly dependent on the axis relative to normal human orientation, so discussions of RF with respect to, e.g., the vertical axis may have only indirect relevance to RF selection on other axes, especially those sagittal or transverse from human orientation.

2 Analysts' Reference Frames

The paper focuses principally on linguistics, but I will briefly characterize the use of the term RF in psychology in so much as it contrasts with the uses in recent linguistic work.

2.1 Psychology

Within psychology, most work on RFs assume a binary distinction between *egocentric* and *allocentric* RFs, the former being determined by the relation from the subject to features in the environment and the latter by arguably more direct spatial relationships present in the environment without regard to the subject. The egocentric RF is well motivated by the sensory system, which necessarily places the subject in the center of the perceived universe. The allocentric RF is largely a negatively defined and mixed collection of spatial categories, which are not dependent on this egocentrism — however much the collection must ultimately derive from sensory input. The collection of allocentric relationships leads to the formation of a *cognitive map* (Tolman 1948), that is, a representation of the environment which stands independently from the immediate subject perspective.

Egocentric mappings are generally thought to be read fairly directly from perception. Less clearly, allocentric mappings may be read from perception when the relevant features to the environment can be directly perceived — as in the typical laboratory context — or they may need to be the result of awareness of non-perceptual information from memory, which in turn may or may not contain egocentric information.

An alternative dichotomy is the *deictic/intrinsic* distinction (see Miller and Johnson-Laird 1976 and Levelt 1984, 1996). Deictic essentially equates with egocentric in that location is determined by reference to asymmetrical features of the subject (or to projections from these features). Intrinsic location is determined by reference to asymmetrical features of a (perceivable) landmark object distinct from the subject. As such, the intrinsic RF is more narrowly defined than allocentric, which in principle need not have any landmark for reference.

The psychological literature is quite broad, but of particular note are more recent neuroimaging studies, which differentially localize spatial knowledge and operations. For example, Aguirre and D'Esposito (1997) report different localizations of landmark appearance and survey (viewer-relative) knowledge using virtual reality and fMRI imaging. Unfortunately route information was not included in this study.

Using ERP data, Taylor et al. (1999) compares viewer-centered and object-centered (intrinsic) interpretations of spatial language with the latter showing a reliably stronger ERP peak and some indication of different localization.

Such studies are difficult to compare not only because of the differing tasks, but the differing types of spatial information queried. Both studies compare conditions with objects features and with orientation, but it is not clear that the same RFs are being evoked in both studies.

Recent discussion of “cognitive maps” in both psychology and geography has developed the notion of alternative perspectives taken by individuals faced with navigational or map tasks. Perhaps the most commonly discussed alternation is

between route and survey perspectives (see especially Taylor and Tversky 1992, Tversky 1993). Route perspectives are essentially the organization of features consistent with actual or imagined travel through a landscape (a building will be to one's left, etc.). Survey perspectives have the god's eye views commonly presented in conventional maps. There is a certain alliance between route perspective and egocentrism (i.e., viewer relative organization). Survey perspectives are allocentric to the extent that they are not organized around the viewer, but clearly calling a survey perspective simply allocentric would leave much missing from a description.

2.2 Linguistics

Linguists have mostly borrowed psychology's ego-/allo- distinction, used "deictic" in a vague way (see the discussion in Levinson 1996), or more recently used a trinary distinction. This trinary distinction largely preserves the egocentric as a type (often called *relative*). The allocentric is divided into an *intrinsic* RF using just features specific to a locating object (often called the *ground*, following Talmy (1978), and an *absolute* RF which uses larger scale co-ordinates (e.g. East) to orient within smaller scale space. For an overview of this style of typology see Pederson (1993), Levinson (1996), Pederson, et al. (1998) and references therein.

The relation between egocentric and relative is problematic. On the one hand, one could simply equate the two, which is sensible under the interpretation that the high salience of the speaker in linguistics is comparable to the high salience of the self in psychology. Under such a characterization, *the cat is at my left* and *the cat is to the left of the trashcan* are both instances of a relative/egocentric RF. Levinson (1996) has argued that this misses the critical distinction of whether or not there is a *transposition* (my term) or "*mapping*" (Levinson's term) of co-ordinates from the viewer/speaker onto the ground. Accordingly, he reserves the use of the term *relative* for only those cases involving transposition. This excludes many expressions using egocentric information from the relative RF. A clearer term might be *transposed relative*.

On such a view, some cases of locating based on the speaker can be seen as a sub-type of intrinsic, namely the use of features of the speaker as object to specify direction. When the speaker is also the locating object ("the cat is to my left"), there is little to distinguish speaker-relative co-ordinates from intrinsic co-ordinates ("the cat is to the car's left"). Both consist of projecting a direction from a feature of a reference object. When there is a ground independent of the speaker ("the cat is to my/the left of the trashcan"), then there is a transposition not found in a simple intrinsic use in which the speaker's orientation is imparted onto the ground object. While distinguishing transposed relative from speaker-as-reference-object is reasonable, the lumping of the speaker with all other reference objects in the intrinsic RF downplays the central importance of the speaker in language production.

Given that the transposed relative RF shares so many features with the intrinsic RF, it is unsurprising that many words and grammatical relations can be used to express either. The intrinsic RF finds expression in all known languages, and the relative RF finds expression in a subset of these. It is likely that the use of terms for the transposed relative RF largely derive historically from (untransposed) intrinsic uses.

We are invited to catalog further special cases, which may also involve transposing co-ordinates from an *origo* (point from which the angular relationships of orientation

is determined) to a ground. The most linguistically obvious would be *altercentric* or *hearer transposed relative* (“the cat is to your left of the trash can”) versus *altercentric intrinsic* (“on your left”). Both are linguistically common for the simple reason of the high computational salience of the addressee in linguistically mediated communication and the awareness that the hearer may not share the speaker’s orientation. In an appropriate context, we might have *third party transposed relative* (“the cat is to his left of the trash can”) and so on.

The division of psychology’s allocentric into intrinsic and absolute is well motivated by the largely distinctive sets of vocabulary used for these frames of reference — at least with respect to the non-vertical axes. Indeed, for roughly planar relations, languages typically use largely the same vocabulary to refer to both relative and intrinsic RFs with frequent ambiguity. On this basis of lexical sets alone, the absolute RF is actually the more distinct with its specialized vocabulary and the relative and intrinsic are more closely related to one another.

This leaves us with a classificational dilemma. Do we wish to have each potentially important contrast as a distinct major branch in the classification schema (giving a potentially n-ary classification)? Do we want to subdivide the intrinsic RF into subtypes, giving us a basic binary distinction between absolute and intrinsic (quite different from psychology’s binary distinction) in which intrinsic subdivides into transposed and non-transposed? Or do we want a trinary subdivision between absolute, non-transposing (formerly known as intrinsic), and transposing (speaker/hearer/etc. relative)?

Levinson (1996), states “As far as we know... there are exactly three frames of reference [intrinsic, relative, and absolute] grammaticalized or lexicalized in language” (138). I would add that the grammatical and lexical contrast between intrinsic and relative seems minimal, so attempts to base this distinction on structural linguistic patterns may prove inadequate.

What then of the absolute RF? Clearly there are subtypes possible for this as well. Despite his clear preference for a three-way typing, Levinson also states “there are reasons for thinking that landmark systems and fixed-bearing systems are distinct conceptual types” (FN35). Indeed, there is something intuitively quite distinct between cardinal direction coordinates (*the cat is north of the trashcan*) and the use of local landmarks (*the cat is towards the wall from the trashcan*). Many language communities fail to support the former absolute strategy while perhaps any language community (with greater or lesser convolution) may support the latter. This suggests that a distinction might be drawn between 1) an absolute RF which relies on co-ordinates which do not derive from the immediate environment and 2) an absolute RF which creates ad hoc co-ordinates by appeal to perceptual features.

Of course, locating by directional reference to perceptual landmarks is akin to locating by projecting from the intrinsic features of a ground object. In both cases there is a direction determined from the ground on the basis of perceptually available features. The difference being that in one case the features belong to the ground and in the other case they do not. It is unclear whether to treat *towards the wall* as a case of transposing co-ordinates onto the trashcan. Subjectively, the example seems to represent a route from trashcan to wall, along which a cat will be found. As such, *the wall* is essentially the same indicator of direction as a more abstract term like *North*.

Obviously a cardinal direction may be ascribed on the basis of perceptual features in the immediate environment, but equally well, the direction may be ascribed on the basis of dead reckoning, which may have a quite indirect relationship to the

environment. Occasionally perceptual and distal landmarks represent an intermediate case.

Problem Cases. The curse of any classificatory system is the cases which defy categorization. In the case of linguistic expression of RFs, many examples suggest a greater flexibility of the assignment of orientation than a simple three way (intrinsic/absolute/relative) distinction.

Consider for example the problem of reference to a complex route description. Directions which are described relative to the route (e.g. *as one heads downtown, the post office will be to the right*) are clearly projections from left/right features. These features could be of an imagined person who is heading in a stated or inferred direction. Alternatively, can we simply say that any route has its own inherent directionality (paralleling *to the snake's left*) and therefore intrinsic and involving imposition of a viewpoint into the scene. Even if we assume that the expression must be for an imagined person rather than for an abstract route, should we then assume that the direction is egocentric specifically to the speaker?¹

What of expressions like *clockwise*? The hands of a clock by convention move clockwise. This rotational direction is presumably intrinsic to the clock, yet the direction of clockwise generally can only be determined by assuming a canonical relationship between the face of the clock and an unspecified observer. The hands of a clock move counterclockwise from the perspective of someone lurking in the back of a grandfather clock — though the mechanism of the clock would certainly disagree with this perspective.

This is similar to *ascribed intrinsic* features such as evidenced by *to the church's left* which is determined by the conventional relationship of the churchgoers to the structure (when on the inside) and by the ascription of a front facet to the entrance of the church from the facing toward the main entrance from the outside. In contrast with ascribed intrinsic, terms like *port/starboard* are used precisely when we want to indicate an intrinsic left/right (of a ship) independent of any human orientation.

However, the term *clockwise* goes beyond the church example in having become a directional term severed from the original referent object (the clock). Now, *clockwise* refers simply to a direction of travel which can only be defined as the opposite of *anticlockwise*. Clockwise travel of an object is not clockwise with respect to another object; it simply is clockwise in and of itself. Clockwise can be defined as the right side of the route being toward the inside of the arc which the route defines, but then clockwise is akin to a route-relative description, as in *the post office will be to the right*. This seems reasonable when clockwise is used with respect to actual travel, e.g. preceding clockwise on an island's coastline.² However, as with *left* and *right* in route descriptions, it is unclear whether to consider *clockwise* as simply an inherently intrinsic feature of any path.

¹ Such route descriptions clearly do not involve a transposition of co-ordinates from the current speaker's orientation to the route unless the route is currently visible to the speaker and hearer. See the discussion of Levelt's deictic vs. intrinsic experiments (Levelt 1984).

² For a description of how terms like "toward the mountain" and "toward the sunrise" shift as one progresses around an island see Wassmann and Dasen (1998). Such is a good candidate for an intermediate case between local landmarks and an absolute co-ordinate system.

Many terms reflect multiple RFs. Familiar examples are terms like *left* and *above* being used both (untransposed) intrinsically and transposed relatively.³ Less discussed are examples which are used in a geographic sense, but variable as to whether they reference local features (environmentally present) or global features (not perceptual in the environment). For example, hill cultures like Tzeltal (Brown and Levinson 1993) or Belhare (Bickel 1997) may use “downhill” for either the local decline or for the more general lay of the land independent of the local decline. Putting aside the question of how these terms are interpreted in context, this alternation is straightforward. However this does suggest that we need to carefully subdivide the absolute RF into at least the subtypes of a) external to the reference, but perceptually available, and b) the more abstract case of external to the reference and overlaid by reference to a global orientation. Some expressions such as *north* may be of only the global type and other especially ad hoc expressions such as *toward the wall* will be of only local. Again, given this distinction, it is less clear that there is a single coherent RF which we can call the absolute.⁴

So thus far, we have found that linguistically both speaker-based terms need to be distinguished according to whether they are transposed or not. Absolute terminology is consistent in assigning direction, but it is highly variable as to whether this is on the basis of features present in the scene. What then of intrinsic? Obviously, intrinsic features may be conventionally assigned (as in the “front of the church” example above). However, this assignment may in its turn rely on features external to the reference object. A trivial case might be the front of a building without distinctive sides being assigned a *front* on the basis of its location to a street. Does the street constitute a local landmark assigning *front* as a facet similar to a north facet being assigned by cardinal directions?

Consider also C. Hill’s (1974) famous rural Hausa example of “front” assignment for a tree being derived from the relationship to a viewer. The side furthest from the speaker/viewer is named the “front”. This is in contrast to the more common English practice (also found in urban Hausa, see Hill 1982) in which the *front* of the tree “faces” the speaker. The English system is inconsistent in that the front of the tree is determined by virtue of the orientation of the tree with respect to the speaker as local landmark and the left/right of the tree is determined by transposing the current viewing orientation onto it. In rural Hausa, a transposed relative system is consistently applied to both front/back and left/right of the tree, i.e., there is a complete transposition of co-ordinates onto the tree. Of course, this is an analysis of the geometrical relations of the speaker, the tree and the shared environment. This analysis cannot be taken as an indicator of the actual cognitive process of facet assignment on the part of the Hausa or English speakers.

Some Tamil speakers may assign a front or a back to a tree on the basis of its relationship within a line of other objects which do have intrinsic fronts and backs. Thus if a tree is in front of a horse by virtue of proximity to the nose of the horse, then the horse may be classed as “behind the tree” on an assumption that the tree is in line

³ See Levelt (1984) for an earlier discussion of this. See Levelt (1996) and Carlson-Radvansky (2000) and references therein for experimental explorations of how this ambiguity is resolved.

⁴ Unlike “downhill” some terms, such as *upwind*, may apply in the absence of perceptual cues (global), but not in contradiction to them.

with the horse and the intrinsic features of the horse determines the orientation of this line (this is described as a type of ascribed intrinsic in Pederson 1993). Importantly, this ascription is independent of speaker/viewer perspective. Since these various ascriptions of intrinsic features rely on culturally and contextually varying calculations, it is also difficult to speak of a unified intrinsic RF across languages.

In addition to having vague boundaries, certain expressions contained within each RF potentially involve calculation which is dependent on features of the current deictic center – e.g., are we inside the church or outside of it? Is it currently windy or not? Accordingly, I do not see an in principle argument for excluding spatial deictic markers which lack angular orientation, such as *here* and *there*, as wholly distinct from RF calculations. Discussions of RFs typically exclude simple proximal/distal deictic markers from consideration because of their lack of angular specification. Alternatively, *here*, *close to*, and *near* can be considered minimal (zero) specifications of angle. Terms like *near*, *behind*, *left of*, *two o'clock from* and *north by northwest of* specify increasingly precise angular relationships.

Many forms which are commonly considered intrinsic also do not specify angular relationships. Consider forms deriving from body parts such as “head” which express “above/over” relationships. In Mixtec (Brugman and Macaulay 1986), these constructions only indicate an “adjacent space” to the “head” in which the figure is to be located. There is no expression of an angular relationship between figure and ground. On the other hand, the expression of a body part of the ground effectively restricts the location to a subset of possible angular relations between figure and ground. That is, the expression has the same effective angular restriction as a more purely angular relationship term such as *above*. We can speak, therefore of a cline from simple expressions of adjacency (*here*) through expressions of part-restricted adjacency (*at the head*) to angular specification (*above*, *in front of*). It is not obvious that there is an exact subportion of this cline which should be deemed intrinsic.

Forms like *near* may also seem distinctive in that they specify relative distance. Forms like *left of* in English ostensibly do not. However, on closer inspection, relative distance is relevant to many expressions of angular specification as well. When the ground object is small relative to the distance from figure, the choice of RF terms becomes restricted. *Left of the church* is only acceptable when the figure is sufficiently proximal. In Mopan Mayan (E. Danziger, p.c.) and Longgu (D. Hill, p.c.), forms translating as “left/right” can only be used for projective space proximal respectively to the viewer’s left/right visual fields. That is, a cup “to the left” and a saucer “to the right” cannot both be to the left of the speaker even if the cup is more to the left than the saucer. In other words, the terms for left/right define regions relative to the speaker in which objects can be located, but they do not define locations of objects relative to one another.

So linguistically speaking, intrinsic/relative/absolute RFs cannot be characterized with clear linguistic expression, nor can boundaries be simply drawn between them. Further, each of these three RFs can be reorganized into less comprehensive, but salient, RFs.

Alternative Categorizations of Reference Frames. Clearly, the grouping of subtypes of RFs varies with the criteria used. Summarizing the above, and elaborating somewhat on Levinson's categorization (Levinson 1996), we can sort RFs according to several major criteria: "Orientation of the figure and ground beyond their internal relationship" is essentially the same as Levinson's constancy of the description "under rotation of whole array".⁵ Ego/allo-centrism is a coarse attempt to relate the principal linguistic RFs to the principal psychology RFs — the importance of the self is clearly important in language as well. Discourse dependence refers to the ability of a description to be correctly interpretable without any contextual knowledge beyond shared knowledge of the ground's location and any general conventions of spatial reference.

Orientation of the figure and ground beyond their internal relationship

Orientation free: Intrinsic

Orientation bound: (Transposed) Relative & Absolute ⁶

Ego/allo-centrism

Egocentric base:

(Transposed) Speaker-Relative

Speaker as ground (special case of intrinsic)

Allocentric base:

Alter- : Hearer-Relative (transposed or intrinsic)

Ground Object-: Intrinsic (except when speaker is ground)

Environmental: Absolute (landmarks, directions, etc.)

Discourse dependence

Discourse dependent:

Speaker/Hearer-Relative, ad hoc Local Landmarks, intrinsic

Discourse independent:

Conventional⁷ and fully distant Landmarks, Cardinal directions

In terms of mapping from linguistic expression to relevant features of RFs, we are left with a largish number of distinct classes: When speakers transpose a projection

⁵ As Levinson notes, one can also sort by properties which are preserved by rotating the viewer/speaker and rotating the ground. I focus on the importance of rotation of the entire array in that whether a relationship is fundamentally autonomous from its environment seems of primary importance. Note, that one could make even finer distinctions such as constancy when relocating viewer to various locations without any rotation, but this would rapidly become cumbersome for no apparent gain.

⁶ This observation was initially developed by Eve Danziger (p.c.).

⁷ Of course, there is a scale from ad hoc to conventional and non-immediate in landmarks. Certainly many abstract indicators of cardinal directions derived historically from conventional landmarks, which presumably conventionalized originally from repeated ad hoc uses. The terms of east and west in Tamil derived over a millennium ago from down(hill) and up(hill), presumably owing to the western highlands and eastern seacoast bounding the Tamil speaking region.

from their own body halves onto an external figure-ground relationship, we can call that a speaker-relative RF. When they do exactly the same set of operations, but using the co-ordinates of an addressee's body, we have a different RF. When they map from their own body onto an object according to a canonical relationship, we have yet another RF. And so on. This expanding list suggests that we would do well to worry less about enumerating RFs than on focusing on the operations underlying such RF assignments. It may be an analytical convenience to characterize speech with pre-packaged RFs. However, speakers are more precisely categorized according to the operations they use: projections from parts, transposition of co-ordinates, assignment of appropriate angular metric, determination of scale, and whatever else which still needs to be determined.

In examining switching from one RF to another, Carlson (1999) describes activation and inhibitory competition between RFs. However, seemingly dramatic changes from one RF to another (e.g. from speaker-relative to absolute co-ordinates) may be the simple result of slightly changing the underlying mix of operations. Under this current view, competition between RFs is better rephrased as the selection of various operations which underlie these RFs. These selections will be conventionally constrained, e.g., a member of a community which never uses cardinal directions in language is unlikely to use cardinal directions for assigning angle. However, within these constraints, a speaker will have far greater flexibility in complex descriptions than simply selecting one of three superordinate RFs. We would do well to record each step of the description in terms of different operations. This finer grained characterization allows tracking more minute adjustments to shifting topic, context, and knowledge.

3 Speaker Selection of Reference Frame Operations

Whether we choose to characterize the shifts a speaker makes as coarse shifts in RF selection or in terms of the underlying operations, the conditions which influence these shifts need to be more comprehensively understood. Experimental tasks can focus on specific shifts from one condition to another, but fall far short of fully accounting for the variation in the wealth of contexts facing speakers every day.

3.1 Alternation in Reference Frames

What is perhaps most remarkable about speakers is not that they have control over a range of RF operations, but that they so effortlessly switch between them. One operation may best suit the micro task of a particular clause and by the next clause, another operation would be the more appropriate. To repeat, this alternation is situated within cultural conventions: if a culture never uses speaker-relative terms for projected space, then we won't expect a speaker within that culture to switch to that system. Further, even cultures which share the same RFs may differ as to when each would be most appropriately used. For example, British and American speakers of English share the left/right/front/back and the NSEW linguistic expressions, but vary in their application of these (Davies and Pederson 2001). Such differences may

partially derive from a differing of the sense of the appropriate scale for the use of each RF.

There may also be a fairly basic difference in how each RF terminology is conceived. Modern Euro-Americans seem locked into a North = Up scheme derived from map conventions of the last 200 years or so. South is the “other way”, east is to the right and west to the left. These terms will be used for intrinsic parts of maps and paper more generally and become an expectation of default orientation of paper representations of space. But even when dealing with actual cardinal directions on the ground, north seems the principal organizing direction in conversation. On the other hand, communities which do not have such map conventions and which extensively use cardinal directions without maps may have no principal orienting direction, or the principal orienting direction is derived suitably enough from prominent environmental features (e.g. the direction of the sunrise).

Charting which RF operations are used in which communicative context is complex and a standard of comparison is necessary for cross-cultural and cross-linguistic comparison. Ideally the standard of comparison would be multidimensional, with each dimension comprising a clear ordinal scale. The sum of the values on all dimensions would correspond to the ideal RF or RFs to use for a description which has these properties. For each speech community, various RFs would be associated with specific ranges on these scales.

3.2 Interactional Scales

There are a number of social conditions (register, gender, expert/novice, task at hand, etc.) which may strongly influence RF selection. Let us focus currently on alternation which is a function of referential conditions. For describing vertical relationships, for example, “absolute” terms are the clear preference in many (and perhaps all) languages — though there may be conditions which favor an intrinsic or egocentric selection.⁸ More subtly, changing referential conditions within a description often trigger a shift from one RF to another, e.g., in a route description previously relying on cardinal directions, the speaker may switch to egocentric terms at the point the traveler nears a local landmark. This commonly corresponds to a switch point from survey to route descriptions in navigational tasks. For example in a route description with a U.S. subject, the dominant cardinal directions are used for the main directions until reaching a local landmark and a decision point:⁹

1) And.. you go uh.. East on fifteenth street? and you'll go past the.. Lane County jail, You go past the.. post office, on your right, and.. you'll go through a.. stop.. uh.. four way intersection stop signs, and.. fifth street will actually.. end. Where you can't go any further, And on that corner, where you can't go straight any longer. you have to turn left or right, ... Yeah the Fifth Street Public Market is.. exactly right there to your right. [Subj. 11, 24/7/00]

⁸ See Carlson-Radvansky (1993) for such an account.

⁹ This task is described in Davies and Pederson (2001).

This switch from using *east* to *to your right* is presumably triggered by the change in geographic scale from block lengths of steady direction to the relatively small size of an intersection and the local turning decision to be made there. See (Montello 1993) for a discussion of the relevance of physical scale to spatial classification.

However, not all alternation will simply correspond to differences in physical size, but may correspond to differences on other scales. Within any given speech community — and holding constant any other parameters which may trigger alternation — the RF is decided from the combined values on these scales.

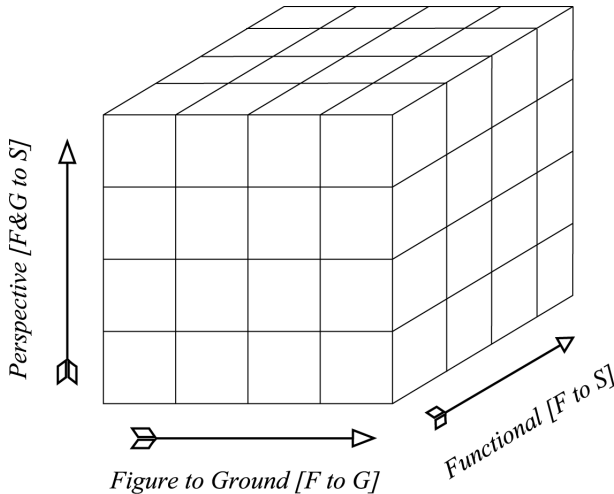


Fig. 1. Interactional scales determining reference frame selection, unexpanded cube

While it remains to be determined which are the most relevant scales for RF alternation, as a starting point, I propose that the validity of this approach be tested with three interacting scales. For convenience these can be represented as a three-dimensional space into which RF usage can be mapped, see Figure 1. For convenience and immediate heuristic purposes, I provide four ordinal values for each scale, but I do not imply that these represent even intervals of discrimination on the scale. Ultimately the number of relevant values will be determined by identifying where on each scale shifts in RF operations are known to occur.

X-axis. “Topological space” scale or the relative geometric relations between Figure and Ground: a) F is a part of G; b) F is in contact with G; c) F lies in an adjacent region of G; d) F lies away from G (e.g., on line from facet)

Y-axis. “Perspective” scale or the degree to which the figure and ground are accessible and/or presupposable within the universe of discourse: a) F & G lie in same topically determined frame or are both in focal attention; b) F & G are both visible or perceptually accessible; c) F is invisible or perceptually inaccessible or perceptually remote; d) F & G are both invisible or perceptually inaccessible or perceptually remote.

Z-axis. “Functional” scale or the relative scale relations between the figure and the discourse participants (especially the Speaker): a) F is body part of S; b) F is

manipulable; c) F and S are interactive (e.g. both people); d) F is geographic (effectively, S can be treated as a point).¹⁰

While I focus on three scales, other scales are potentially relevant. Topicality/familiarity (the role of the ground in the current discourse) would be a good candidate for a fourth axis.

The configuration or shape of the ground object is quite relevant for the determination of the most appropriate preposition (or preposition equivalent), see especially Landau and Jackendoff (1993), but does not seem particularly important for RF discrimination. An exception would be the necessity for the ground to have the appropriate parts to sanction use of certain intrinsic descriptions. For example, one can only sit “at the head” when a table can be deemed to have such a part. In the absence of such an obvious part, a speaker might select different co-ordinates. This is not a particularly scalar notion, but rather a simple requirement of any feature assignment. Another relevant, but non-scalar, factor is whether or not speakers choose to use their own bodies as ground object (“the box is on my left/in front of me”). Presumably such a decision is based on the adequacy of speaker-as-ground for specifying location in the current discourse. The determination of this adequacy, I leave for future study.

The sum of these three selected scales of four values creates a three dimensional matrix (or *cube*) of 64 cells. Many cells clearly refer to improbable referential situations, but for the rest, the researcher can determine which RF operations are most likely to be used for descriptions of that situation — holding social conditions, etc. constant. This matrix is shown in exploded form in Figure 2.

To the extent that each operation is used only for contiguous cell values, we can say that the scales are relevant to the determination of RF. If certain RF operations are used for noncontiguous cells — that is, expression of an intermediate cell with that RF is impossible — then the hypothesized scales have been falsified.

This method also allows us to identify any interaction among the scales in determining the RF operations. That is, what might appear to be a purely holistic effect of RF alternation occurring across complex referential alternations can be decomposed into distinguishable values on different scales.

Alternations in RF operations can be mapped for an individual’s language performance. However, this approach is particularly well suited for mapping consistent patterns within a speech community. Once multiple speech communities have been mapped in this way, precise and uniform comparison across speech communities finally becomes possible.

Since a number of these cells contain references which are quite unlikely to be spoken of, full consideration of this initial list of 64 cells is unnecessary. Other cells seem likely to reveal different patterns from their neighboring cells or have had cross-linguistic variation reported. For explication, I list the cells I judge to usable in normal conversation with invented English examples in Table 1 (in two parts).

¹⁰ Montello (1993) appropriately subdivides geographic scale into smaller divisions which clearly trigger different linguistic calculations.

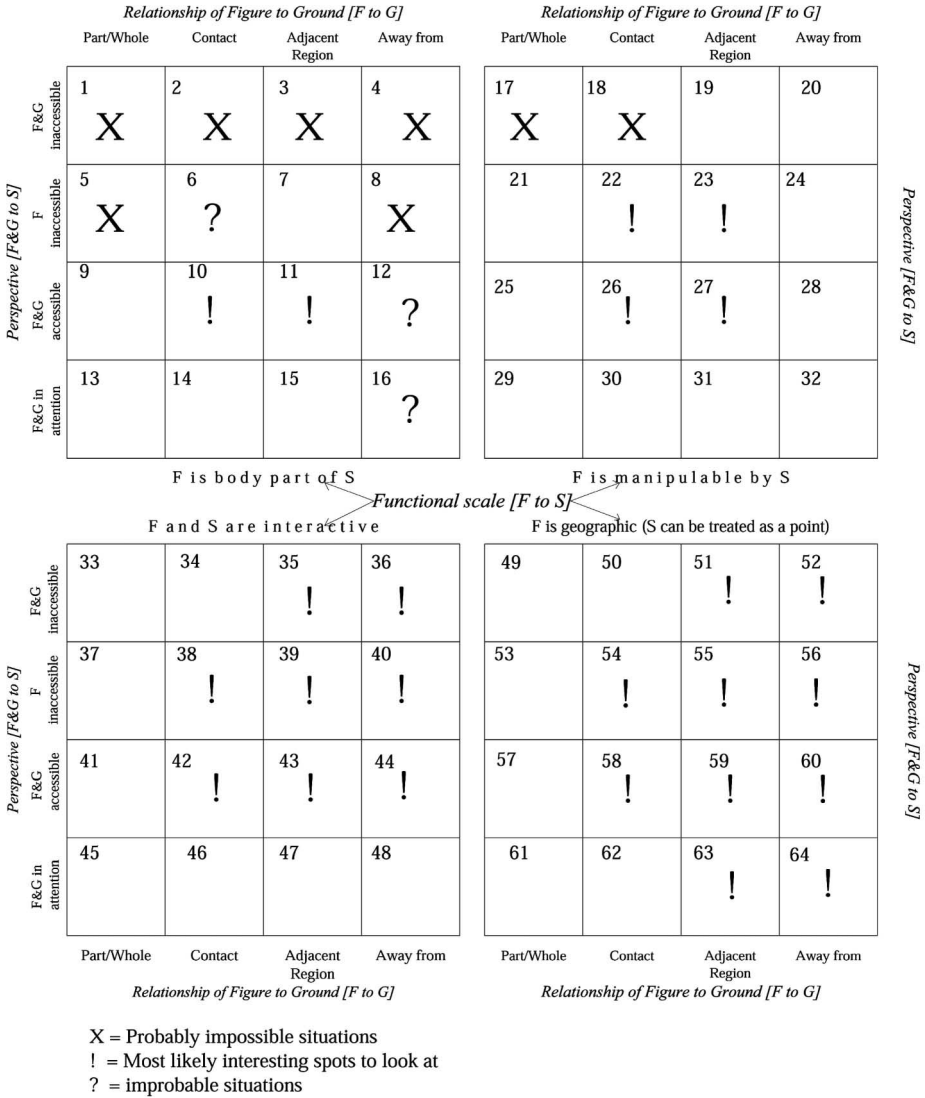


Fig. 2. Interactional scales determining reference frame selection, expanded

After filling in the cells for multiple speech communities, we will also be able to determine any implicational relations among the values in this three dimensional representation. For example, it may prove to be an implicational universal (in the sense of Greenberg 1978) that if a given language allows the use of a cardinal direction term for cell 32, then that language will allow use of a cardinal direction term for cells 48, 60, 63, and 64. In this way, different languages with importantly diverse patterns of RF use can nonetheless be related as coherent subtypes within a single (even universalist) account.

Table 1 (1st part). Individual cell values for the interactional scales

Cell	F to G	Vis to S	F funct to S	Example (Ground)
?6	contact	G only	part of S	<i>The mole is in the left area of <u>my armpit</u></i>
7	adjac	G only	part of S	<i>The mole is left of <u>my shoulder blade</u></i>
9	part	F&G	part of S	<i>My broken toe is on the left side of <u>my foot</u></i>
!10	contact	F&G	part of S	<i>My leg is against the (left of the) <u>chair</u></i>
!11	adjac	F&G	part of S	<i>My sore elbow is left of my <u>chest</u></i>
?12	away	F&G	part of S	<i>My head is lying due east of <u>the streetlamp</u></i>
13	part	FG attn	part of S	<i>The broken toe is one the left of <u>this foot</u></i>
14	contact	FG attn	part of S	<i>My hand is on the left of <u>the bucket</u></i>
15	adjac	FG attn	part of S	<i>My head is to the left of <u>your pillow</u></i>
?16	away	FG attn	part of S	<i>My head is in the shadow of <u>this post</u></i>
19	adjac	Neither	manipulable	<i>My key should be near <u>where I lost it</u></i>
20	away	Neither	manipulable	<i>The ball should be ahead of <u>where it was kicked</u></i>
21	part	G only	manipulable	<i>The handle must be behind <u>the pot</u></i>
!22	contact	G only	manipulable	<i>The lid must be against the back of <u>the pot</u></i>
!23	adjac	G only	manipulable	<i>The cup must be over to the left of <u>the pot</u></i>
24	away	G only	manipulable	<i>The cup must be out left from <u>the pot</u></i>
25	part	F&G	manipulable	<i>The handle is on the left of <u>the pot</u></i>
!26	contact	F&G	manipulable	<i>The broom is leaning on the left of <u>the box</u></i>
!27	adjac	F&G	manipulable	<i>The cup is standing to the left of <u>the pot</u></i>
28	away	F&G	manipulable	<i>The box is 6m. toward the door from <u>the fridge</u></i>
29	part	FG attn	manipulable	<i>The handle is on the left of <u>this pot</u></i>
30	contact	FG attn	manipulable	<i>That broom is leaning on the left of <u>this bucket</u></i>
31	adjac	FG attn	manipulable	<i>That box is near the left of <u>this pot</u></i>
32	away	FG attn	manipulable	<i>That box is about 6m. left of <u>this pot</u></i>
33	part	Neither	interactive	<i>The driver is on the right (of <u>car</u>) in England</i>
34	contact	Neither	interactive	<i>Bill is probably hiding behind <u>his closet door</u></i>
!35	adjac	Neither	interactive	<i>Bill will be waiting in front of <u>his door</u></i>
!36	away	Neither	interactive	<i>Bill should be parked north of <u>the church</u></i>
37	part	G only	interactive	<i>That stained glass should be on <u>the back wall</u></i>
!38	contact	G only	interactive	<i>Bill should be at <u>the back door</u></i>
!39	adjac	G only	interactive	<i>Bill should be behind <u>this church here</u></i>
!40	away	G only	interactive	<i>Bill should be somewhere north of <u>this point</u></i>
41	part	F&G	interactive	<i>Bill is on the left side of <u>the congregation</u></i>
!42	contact	F&G	interactive	<i>Bill is leaning against the left of <u>his desk</u></i>
!43	adjac	F&G	interactive	<i>Bill is standing to the left of <u>the puddle</u></i>
!44	away	F&G	interactive	<i>Bill is out in back of <u>the church</u></i>
45	part	FG attn	interactive	<i>This guy is in among the front of <u>the group</u></i>
46	contact	FG attn	interactive	<i>This guy is leaning against <u>his desk</u></i>
47	adjac	FG attn	interactive	<i>This guy is standing to the left of <u>his desk</u></i>
48	Away	FG attn	interactive	<i>This guy is standing back from <u>his desk</u></i>

Table 1 (Continued). Individual cell values for the interactional scales

Cell	F to G	Vis to S	F funct to S	Example (Ground)
49	part	Neither	geographic	<i>Amsterdam is at the top of <u>the Randstad</u></i>
50	contact	Neither	geographic	<i>Poland is east of <u>Germany</u></i>
!51	adjac	Neither	geographic	<i>Arnhem is north of <u>the Waal</u></i>
!52	away	Neither	geographic	<i>Oslo is quite a bit north of <u>Amsterdam</u></i>
53	part	G only	geographic	<i>Eugene is in the west of <u>Oregon</u></i>
!54	contact	G only	geographic	<i>The industry borders to the south (of <u>here</u>)</i>
!55	adjac	G only	geographic	<i>Springfield is over to <u>our</u> east</i>
!56	away	G only	geographic	<i>The Cascades run to the east of <u>here</u></i>
57	part	F&G	geographic	<i>The Psych bldg is on the east edge of <u>campus</u></i>
!58	contact	F&G	geographic	<i>The mall area lies just above <u>downtown</u></i>
!59	adjac	F&G	geographic	<i>The hotel comes on the left after <u>the mall</u></i>
!60	away	F&G	geographic	<i>Downtown is west of the <u>university</u></i>
61	part	FG attn	geographic	<i>Our home is in the South Hills of <u>Eugene</u></i>
62	contact	FG attn	geographic	<i>That parking lot is right behind <u>our building</u></i>
!63	adjac	FG attn	geographic	<i>That street is due south from <u>our building</u></i>
!64	away	FG attn	geographic	<i>From here, Eugene lies in a line with <u>Salem</u></i>

This representation can also determine which cells, if any, recurrently motivate switches of RF operations, independently from which RFs happen to be relevant for particular speech communities. Such “critical cells” would suggest which referential domains call for more detailed semantic investigation. An understanding of the relevant semantic parameters as well as the implicational relations operating across them could improve our understanding of RF alternations.

The single example for each cell is not intended to suggest that there will be a single type of expression for each set of values. Rather there should be at least one preferred mode of expression and possibly a few less typical expressions. Importantly, perhaps every cell will have at least one prohibited mode of expression. Each of these should be collected within a speech community for each cell. For simplicity’s sake, this discussion ignores variation which is the result of style and speech context.

3.3 Using such a Matrix — A “Wish List”

Ultimately, if we chart RF use as largely derivative from the values on the relevant scales, then we should have a tool for

- Developing precise models of ontogenetic development. Do RF operations “spread” from an initial use in just a few cells to larger collections of cells?
- Creating predictions of processing times and error patterns in locational tasks. Certain RF operations can be predicted to be inherently more difficult for certain cell values.

- Explaining competition and resolution between alternate locational characterizing. Does a choice of RF operations for a given cell reflect that the cell is on the boundary between two sets of cells each of which has a dominant expression?
- Explaining the effects of context (linguistic, referential, experiential, etc.) on selection. Do certain contextual features make the values on some scales particularly salient? Alternatively, do some contexts effectively neutralize any contribution of some factors? E.g. could physical scale become irrelevant in a discussion of geometrical properties?
- Determining how cell values need to be calculated differently depending on the axis on which a spatial discrimination is being made (sagittal vs. transverse vs. vertical in relation to the speaker/hearer).
- Explaining training and learning of novel RF uses (including situations of cultural contact)

3.4 Cultural Shifts

To address the last item on the above “wish list”, let us ask what motivates a shift within a cultural group. Consider the case of relative vs. absolute navigation. Both are demonstrably adequate for navigation in a wide range of cultures, so we should be skeptical of attempts to relate the cultural preferences of particular RFs to simple shifts in geographic living conditions. At least in some cases, a shift in the dominant RF for a given situation may be the direct result of lexical borrowing from language contact.

For example, during the 1990’s I worked with some members of the Bettu Kurumba, who are a traditionally hunter-gatherer society in the Nilgiri foothills in South India. Over the past few decades, the Bettu Kurumba have been resettled into camps under direct (Tamil) state administrative control. Traditionally, the Bettu Kurumba made extensive use of local landmarks for navigation with perhaps occasional egocentric reference. For example, there are no native terms for cardinal directions. In descriptions of manipulable space, local landmarks were less used.

In contrast to the traditional Bettu Kurumba, the surrounding rural Tamil culture typically uses cardinal direction terms for both geographically scaled space and for locations of manipulable objects. Since the 1980’s, the majority of Bettu Kurumba children now attend Tamil-medium schools and bilingualism with Tamil has become standard in many Bettu Kurumba communities. The Tamil words for “north”, etc. have been borrowed into the school children’s Bettu Kurumba. However, this cannot be a simple lexical borrowing, for a novel system of calculations using cardinal directions must be borrowed along with the lexical items. By 1992, children as young as about seven were spontaneously describing even manipulable space using cardinal directions. (E.g. “put the [toy] pig north of the [toy] cow”.)

Such examples suggest that perhaps RF selection is in large part lexically driven. Without the relevant vocabulary at hand, a RF will not be used. As exposure to RF-specific vocabulary increases, the appropriateness of the use of operations associated with that RF increases correspondingly. Rather than trivializing a complex process, the task of lexical acquisition needs to be understood as a complex process of developing new cognitive operations.

4 Summary

While it is clear that each discipline concerned with spatial orientation will continue to use notions of RFs in somewhat discipline-specific ways, the linguistic data suggest that coarse-grained categorization of RFs could be refined into an understanding of smaller more specific operations. Since speakers so readily shift from one operation to another as referential content shifts, the analyst needs some method to track these shifts. With a well-delimited, multidimensional model of the relevant parameters, it should prove possible to establish standards of comparison for individual, linguistic, and cross-cultural variation in patterns of RF use.

References

- Aguirre, G.K. and D'Esposito, M.: Environmental knowledge is subserved by separable dorsal/ventral neural areas. *Journal of Neuroscience* 17 (1997) 2512-2518
- Bickel, B.: Spatial operations in deixis, cognition, and culture: where to orient oneself in Belhare. In Nuyts, J. and Pederson, E.(eds.): *Language and conceptualization*. Cambridge University Press, Cambridge (1997) 46-83
- Brown, P. and Levinson, S.C.: "Uphill" and "Downhill" in Tzeltal. *Journal of Linguistic Anthropology* 3 (1993) 46-74
- Brugman, C. and Macaulay, M.: Interacting semantic systems: Mixtec expressions of location. In Nikiforidou, V., VanClay, M., Niepokuj, M. and Feder, D.(eds.): *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society February 15-17, 1986*. Berkeley Linguistics Society, Berkeley (1986) 315-327
- Carlson, L.A.: Selecting a reference frame. *Spatial Cognition & Computation* 1 (1999) 365-379
- Carlson-Radvansky, L.A.I., David E.: Frames of reference in vision and language: Where is above? *Cognition* 46 (1993) 223-244
- Carlson-Radvansky, L.A.T., Zhihua: Functional influences on orienting a reference frame. *Memory & Cognition* 28 (2000) 812-820
- Davies, C. and Pederson, E.: Grid patterns and cultural expectations in urban wayfinding. In Montello, D.R.(ed.) *Spatial Information Theory*. Springer-Verlag, Berlin (2001) 400-414
- Greenberg, J.H.: Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, J.H.(ed.) *Universals of language*. Readings of linguistics. MIT Press, Cambridge (1978) 73-113
- Hill, C.: Spatial perception and linguistic encoding: A case study in Hausa and English. *Studies in African Linguistics* 5 (1974) 135-148
- Hill, C.: Up/down, front/back, left/right: A contrastive study of Hausa and English. *Pragmatics and Beyond* 3 (1982) 13-42
- Landau, B. and Jackendoff, R.: "What" and "where" in spatial language and spatial cognition. In *Behavioral and Brain Sciences* 16 (1993) 217-238
- Levelt, W.J.M.: Some perceptual limitation on talking about space. In van Doorn, A.J., van de Grind, W.A. and Koenderink, J.J.(eds.): *Limits in perception: essays in honour of Maarten A. Bouman*. VNU Science Press, Utrecht, The Netherlands (1984) 323-358
- Levelt, W.J.M.: Perspective taking and ellipsis in spatial descriptions. In Bloom, P., Peterson, M., Nadel, L. and Garrett, M.(eds.): *Language and Space*. MIT Press, Cambridge, Massachusetts (1996) 77-107
- Levinson, S.C.: Frames of References and Molyneux's Question: Crosslinguistic Evidence. In Bloom, P., Peterson, M., Nadel, L. and Garrett, M.(eds.): *Language and Space*. MIT Press, Cambridge, Massachusetts (1996) 109-169

- Miller, G.A. and Johnson-Laird, P.N.: Language and perception. Belknap Press of Harvard University Press, Cambridge, Massachusetts (1976)
- Montello, D.R.: Scale and multiple psychologies of space. In Frank, A.U. and Campari, I.(eds.): Spatial Information Theory. Springer-Verlag, Berlin (1993) 312-321
- Pederson, E.: Geographic and manipulable space in two Tamil linguistic systems. In Frank, A.U. and Campari, I.(eds.): Spatial Information Theory. Springer-Verlag, Berlin (1993) 294-311
- Pederson, E., Danziger, E., Levinson, S., Kita, S., Senft, G. and Wilkins, D.: Semantic typology and spatial conceptualization. *Language* 74 (1998) 557-589
- Talmy, L.: Figure and ground in complex sentences. In Greenberg, J.H.(ed.) *Universals of human language*. Stanford University Press, Stanford, California (1978) 625-649
- Taylor, H.A. and Tversky, B.: Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language* 31 (1992) 261-282
- Taylor, H.A.N., Susan J; Faust, Robert R; Holcomb, Phillip J.: "Could you hand me those keys on the right?" Disentangling spatial reference frames using different methodologies. *Spatial Cognition and Computation* 1 (1999) 381-397
- Tolman, E.C.: Cognitive maps in rats and men. *Psychological Review* 55 (1948) 189-208
- Tversky, B.: Cognitive maps, cognitive collages, and spatial mental models. In Frank, A.U. and Campari, I.(eds.): Spatial Information Theory. Springer-Verlag, Berlin (1993) 14-24
- Wassmann, J. and Dasen, P.R.: Balinese spatial orientation: some empirical evidence for moderate linguistic relativity. *Journal of the Royal Anthropological Institute (New Series)* 4 (1998) 689-711

Motion Shapes: Empirical Studies and Neural Modeling

Florian Röhrbein¹, Kerstin Schill¹, Volker Baier², Klaus Stein²,
Christoph Zetzsche¹, and Wilfried Brauer²

¹ Institut für Medizinische Psychologie,

Ludwig-Maximilians-Universität München, Germany

² Institut für Informatik, Technische Universität München, Germany

Abstract. Any mobile agent able to interact with moving objects or other mobile agents requires the ability to process motion shapes. The human visual system is an excellent, fast and proven machinery for dealing with such information. In order to obtain insight into the properties of this biological machine and to transfer it to artificial agents we analyze the limitations and capabilities of human perception of motion shapes. Here we present new empirical results on the classification, extrapolation and prediction of motion shape with varying degrees of complexity. In addition, results on the processing of multisensory spatio-temporal information will be presented. We make use of our earlier argument for the existence of a spatio-temporal memory in early vision and use the basic properties of this structure in the first layer of a neural network model. We discuss major architectural features of this network, which is based on Kohonens self-organizing maps. This network can be used as an interface to further representational stage on which motion vectors are implemented in a qualitative way. Both components of this hybrid model are constrained by the results gained in the psychophysical experiments.

1 Introduction

What are the motion primitives or prototypical motion shapes which are used by the visual system in order to classify and predict trajectories? This question guided the experiments described in the subsequent sections. We applied a number of different experimental paradigms in which we increased successively the complexity of the motion stimuli used. We started with simple kinks and curves (section 2.1), went further with occluded paths (section 2.2), multimodal motion stimuli (section 2.3) and ambiguous displays (section 2.4), and ended with extended and very complex trajectories (section 2.5). In all experiments we varied temporal parameters in order to gain insights into memory-based processes. Since in natural situations the behavior of the biological system is influenced by more than one sensory system we also present results in which we investigated how humans process spatio-temporal information from different sensory systems.

The empirical results on the motion shape vocabulary of the visual system are transferred to our modeling approaches. These approaches include the consideration of different levels of processing and representation. For the higher, more cognitive level of spatio-temporal information processing a propositional framework for the

qualitative representation of motion information has been developed. It uses linguistic concepts for qualitative motion vectors like left-turn, U-turn, loop etc. which are correlated to motion primitives found in our experiments. This qualitative approach can be used to describe, generalize and calculate trajectories on different levels of abstraction (described in detail in [9]). However, there remains a gap between a first early visual stage of dynamic processing (as provided by an orthogonal memory stage as suggested in [16]) and a linguistically coded qualitative representation of motion information on the other side.

Therefore we started to investigate multi-level neural architectures, which are suitable for multi-stage processing of spatio-temporal information. In order to enable the representation, processing and prediction of spatio-temporal patterns on different levels of granularity, a hierarchical network model has been investigated which consists of Kohonens self-organizing maps [6] organized in a hierarchical manner. The model has the advantage of a self-teaching learning algorithm and stores temporal information by local feedback in each computational layer. Its basic architectural properties are described in section 3.

2 Experimental Results

In all experiments described below the subjects were sitting in a semi-darkened room in front of a computer display at a viewing distance of about one meter. As stimulus they saw a black dot moving on a white screen along an invisible path (except the multimodal experiment described in 2.3). Every session lasted about 50 minutes and the subjects (most of them medical students, males and females) were paid for their participation.

2.1 Discrimination of Changing Direction

In the first series of experiments we measured the ability of the visual system to discriminate simple motion trajectories. The dynamic stimulus consisted of a black dot (0.1 deg vis) moving with constant speed along an invisible angled path (see Fig. 1). In each trial a pair of such trajectories was presented subsequently with an ISI of 300 msec. The subjects' task was to decide whether there was a greater change in the direction of motion in the first or in the second stimulus and to indicate this by pressing one of two buttons. The stimuli were arbitrarily rotated (also within trials) to prevent the subject from using additional cues, like external reference points.

Presentation time (i.e. the duration of dot movement) was varied in 17 conditions from 100 to 2500 msec, the difference between reference and test stimulus was varied in three spatial conditions (Fig. 1, middle). In addition, we modified the turnaround by replacing the described motion pattern (trajectory with a kink) with a stimulus consisting of a curved trajectory. These stimuli differed in the degree of curvature and the corresponding task was to determine which trajectory was more curved.

We had five subjects, all of them attended five sessions and in each session all 3*17 conditions were measured ten times.

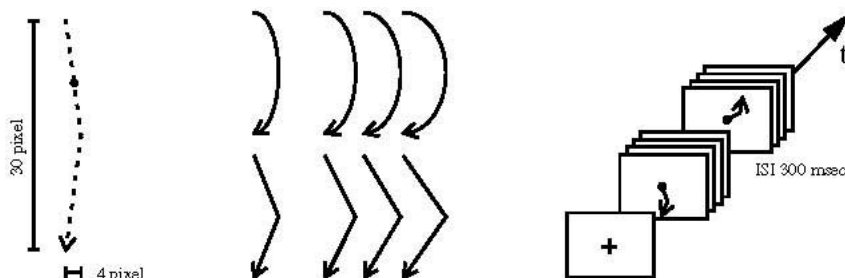


Fig. 1. Left: The trajectory of the dynamic reference stimulus is four pixel (0.08 deg vis) in width and 30 pixel (0.7 deg vis) in length. Middle: Stimulus pairs used for ‘continuous’ and ‘discrete’ mode (first resp. second row). The test stimuli are five, six or seven pixel in width, which lead to small, intermediate or large spatial differences. Right: Time course of a single trial in the 2AFC design used.

Results. In all conditions the main result was a slight increase in performance up to several hundred msec stimulus duration, followed by a widely invariant performance up to 2.5 sec. This is surprising since the visual impression of a motion that lasts of only a few hundred msec is very different from one lasting 2.5 seconds, a result which has to be taken into account in our modeling approach. The probability of correct decisions was in the desired range between nearly chance and almost perfect performance. As expected, the more the trajectories differed in curvature the better the performance was. A less accurate processing with continuously varying direction of motion (Fig. 3) could be observed in comparison to trajectories with an abrupt change in direction (Fig. 2). This task-dependent accuracy defines appropriate parameters values and constrains for our neural network model.

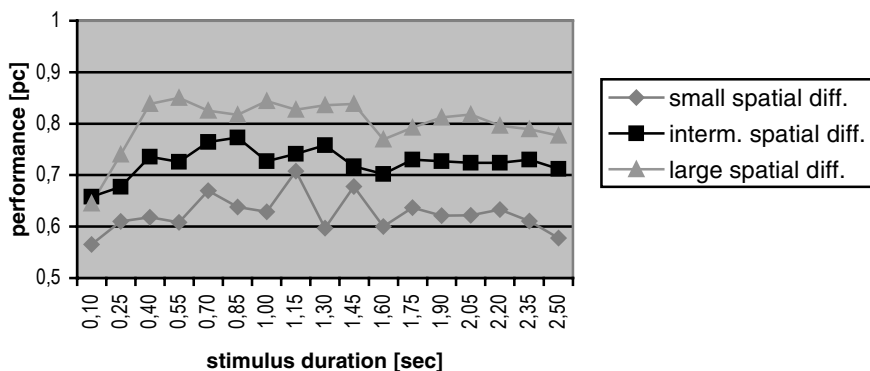


Fig. 2. Resulting performance (percent correct) for straight dot movements containing a kink for different spatial conditions (‘discrete’ mode).

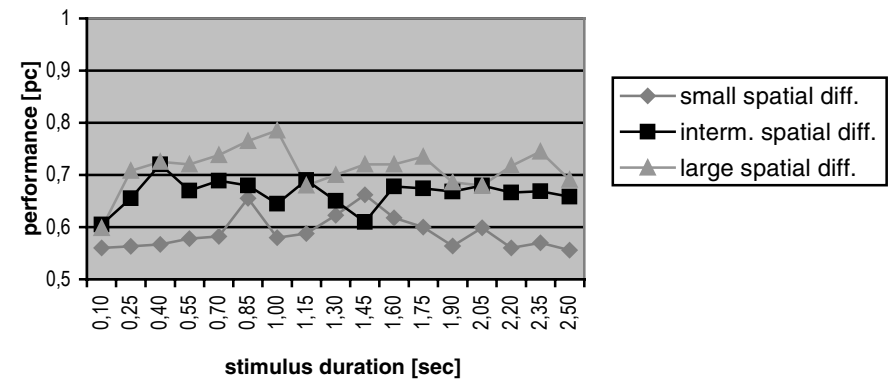


Fig. 3. Resulting performance (percent correct) for curved trajectories (‘continuous’ mode). For the three different conditions see Fig. 1.

Discussion. Neither relatively fast movements nor long durations have an influence on subjects’ sensitivity. This invariance is supported by previous experiments on direction and curvature discrimination [4]. These experiments with curved motion paths also revealed that the discrimination performance was significantly different (better) when corresponding static versions of the dynamic stimuli were used. Therefore the observed invariance can neither be explained by motion-selective mechanisms nor by the assumption of static internal representations in the form of spatial trajectories of the dynamic stimuli.

2.2 Prediction of Spatio-Temporal Pattern

In order to understand more about the processes involved in the prediction of motion paths we conducted several experiments adapted from studies on the Poggendorff illusion (see e.g. [18]). The classical static version of the Poggendorff display consists of a vertical bar intersected by two oblique lines lying along the same straight line. The illusion is derived from the observation that most people report the two portions of the transversal to be displaced from true geometric alignment.

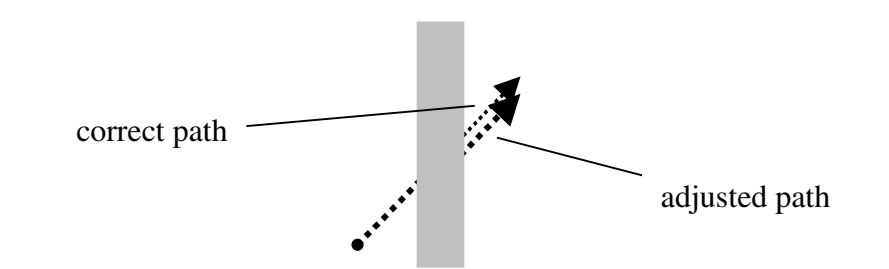


Fig. 4. The dynamic version of the Poggendorff display.

In the dynamic version (Fig. 4) subjects saw a point moving along this (invisible) transversal (cp. [10]). Their task was to adjust the position on the right side of the occluding bar where the moving dot will reappear. For this they observed the moving dot several times (usually four to six times) and adjusted the position after each trial before they pressed a key in order to proceed with the next presentation.

We had conducted some first experiments, which revealed that the prediction accuracy becomes worse when the occluding bar is wider and the dot velocity slower (see [4]). In the experiments there dot velocity was kept constant in order to invoke the impression of a homogenous movement, but this resulted in a mixing of several spatial and temporal effects.

Experiment 1. In the study here we relaxed the restriction of a homogenous movement and varied the dot velocity separately for the two visible parts and for the temporal gap between them (i.e. the occlusion time). The dot velocity was 5 or 15 deg/sec, which resulted in a presentation time of 400 resp. 130 msec for each visible part. The occlusion time could be 0, equal the dot velocity (1200 resp. 400 msec) or it could be twice as long (2400 resp. 800 msec). Additionally we devised a static-alternating version where the whole trajectory of the left portion was presented at once and after the occlusion time had passed, the whole right portion was presented at once. We measured the alignment error (AE), whereby a positive AE means that the adjusted segment was lower than the correct one (as in Fig. 4).

The results for 18 subjects are depicted in Fig. 5. They show that presentation mode and both temporal parameters tend to influence the accuracy. We found that decreasing the presentation time (increasing the dot velocity) and decreasing the occlusion time (temporal gap) often lead to smaller alignment errors. A static presentation always leads to a more precise alignment than a dynamic one.

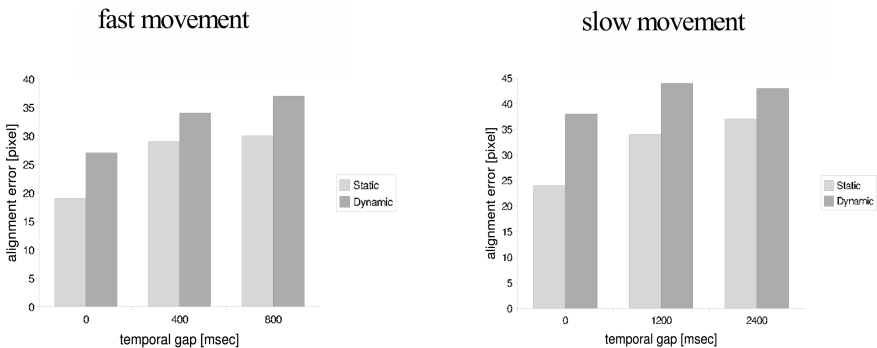


Fig. 5. Systematic misalignment for different velocities, temporal gaps and presentation modes.

Experiment 2. To get more detailed and reliable information about the influence of the temporal gap, in a second experiment, velocity was kept constant and the occlusion time was varied in 11 conditions from 0 to 2.5 sec in steps of 250 msec. The bar width could be small (80 pixel, 2.4 deg vis) or large (115 pixel, 3.4 deg vis). The results in Fig. 6, again, show a very systematic alignment error. This error is positive, i.e. the hidden path is always underestimated and the error is more pronounced for

larger spatial gaps. However, for both bar widths the subjects show a strikingly constant performance with respect to the occlusion time.

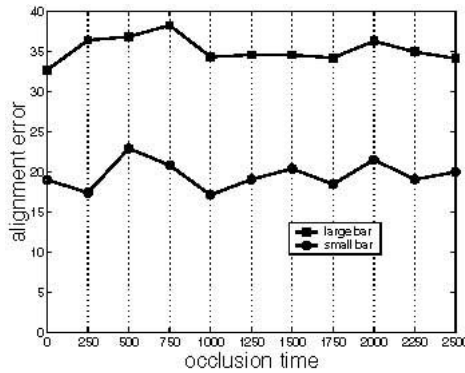


Fig. 6. Results for ten subjects and eight repetitions each in all 22 conditions.

Discussion. The strength of the illusion therefore seems not to depend on the temporal gap, but only on the spatial one. In a related study [11] the subjects had to estimate the time when a moving target would have passed a certain spatial position. Their results show that the performance in this temporal task is independent of spatial parameters, but only temporal parameters had an influence. So one might speculate that there are task-dependent representations for the processing of spatio-temporal pattern and that the system has the capability to switch from a spatial strategy to a temporal one.

These results will influence the benchmarks for the prediction quality of the neural network model. As an extension of this visual prediction task we are planning an experiment with multisensory stimuli by adding an acoustic signal to a moving spot of light. We expect insights in the underlying integration mechanisms, especially in the systems capability of improving prediction accuracy by exploiting redundant information.

2.3 Audio-visual Interactions

Here we have extended our studies on visual motion processing by employing auditory-visual stimuli. The usage of multimodal stimuli can be regarded as one step towards a more natural stimulation, since under normal conditions several modalities are stimulated at the same time by one object or one event. This is especially true for the observation of a moving sound-emitting object like an animal or vehicle. If observer and object come closer or move apart, the resulting stimulation causes systematic covariations in the visual and auditory channels. An unresolved question is how this redundant information is pooled. The quite limited knowledge about multimodal perception prompted us to investigate the perceptual performance in a basic stimulus configuration. This work was done in an associated project¹.

¹ DFG graduate college GRK 267 Sensorische Interaktionen

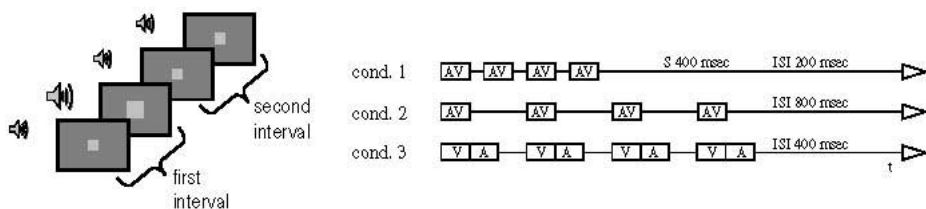


Fig. 7. Stimuli (left) and time course (right) of auditory-visual compounds (three conditions).

We presented two light/tone combinations in a first interval, and after a short pause another two combinations in a second interval. The auditory component consisted of a 1kHz tone and was presented via headphones. The visual stimulus was a gray square of 2.5 deg vis. For one pair we varied the intensity of the auditory and/or the size of the visual stimulus component by adding small in- or decrements to the reference compound, such that the subjects can have the impression of a sound-emitting object which comes closer or moves away in space. The subjects' task was to detect any change in the bimodal stimulus configuration independent of modality. They responded by pressing one of two buttons (without time constraint).

For these bisensory intensity/size differences we measured the two-dimensional threshold curve by determining eight combined auditory-visual difference thresholds. These thresholds were determined with an adaptive procedure, which lead to about 40 presentations per threshold. Within one session four thresholds were measured in carefully selected combinations in order to ascertain that the subjects attend to both signal components all the time.

There were three conditions: Presentation time was 400 msec, with an inter stimulus interval (ISI) of 200 or 800 msec (condition 1 resp. condition 2). In a third condition we suspended the temporal coincidence by sequential presentation (visual component 400 msec, auditory component 400 msec, ISI 400 msec).

Results. In Fig. 8 results are plotted as difference threshold curves around the reference compound at the axis origin. Positive values indicate increments, negative values decrements. Points on the jnd-curve represent measured relative thresholds averaged over subjects. Since we are interested in the relative performance the data were scaled to the four unimodal thresholds on the abscissa and ordinate.

Discussion. The subjects show substantial intersensory facilitation effects. As can be seen from the shaped regions the obtained redundancy gain cannot be explained as a mere statistical summation effect [13]. Moreover, the amount of facilitation seems to depend on the sign of the auditory and visual variation. In particular the difference thresholds for the ecologically relevant quadrants with simultaneous increments or decrements for light and tone show a significant increase in sensitivity. This is evidence for a true bisensory summation [5, 8] of the incoming signals at a relatively early stage in the system.

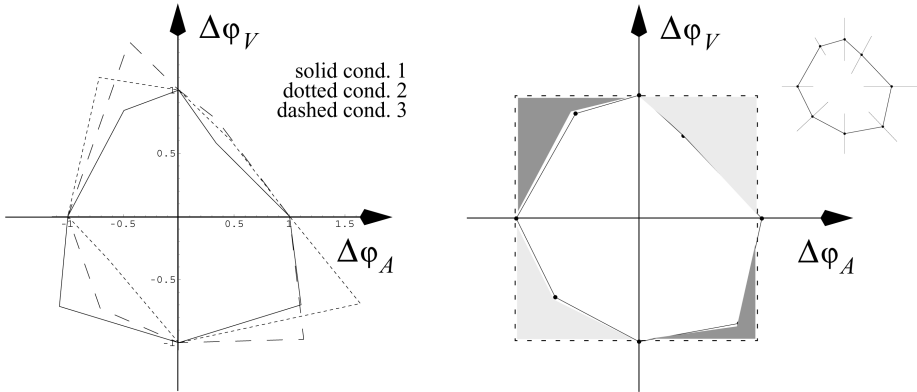


Fig. 8. Measured difference thresholds with respect to auditory variation (on the abscissa) and visual variation (on the ordinate). Left: Results for two subjects in three conditions. Right: Averaged performance with error bars plotted in the inset.

2.4 Extrapolation of Motion Paths

With more extended (unimodal) stimuli we designed an experiment conceptualized to uncover the mechanisms involved in the prediction and extrapolation of visual motion information. With an ambiguous stimulus setting we investigated the influence of the history of trajectories on parameters of a spatio-temporal memory structure. These parameters are sought to determine the perceived motion path.

The apparent motion display used is shown in Fig. 9., numbers indicate frames at successive points of time. Two dots are presented simultaneously and the subjects see two movements, one upward movement from down left and one downward movement from top right. The end of the motion paths can thereby be perceived in two different ways, either curved or straight, depending on the assignment of the last frame (number 4 in Fig. 9).

It is well known that for such stimulus configurations the perceived motion depends on the distance of the points (rule of proximity) and on the history of the path [14]. Since in our display all points are equidistant, the percept should critically depend on the past trajectory. In the experiment we thus varied the history of the path (curved or straight), their length, i.e. the number of frames (3,5,7) and additionally the SOA (150, 200, 250 msec). The stimulus example in Fig. 9 shows two trajectories which have a curved history consisting of four frames each. In all conditions, the subjects had to indicate which kind of movement they had perceived, additionally reaction times were recorded.

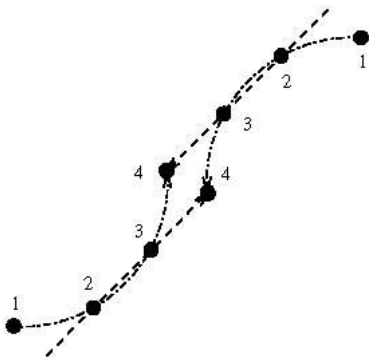


Fig. 9. Apparent motion display with two movements, perceivable in two ways (dashed lines).

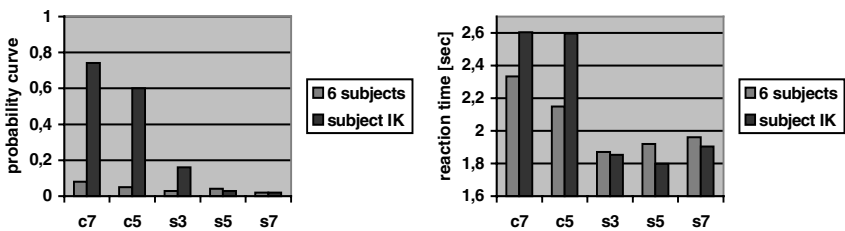


Fig. 10. Probability of perceiving a curve and corresponding reactions times for five conditions. Results are pooled over three temporal conditions and six repetitions per subject.

Results. The subjects' answers are shown on the left side of Fig. 10. On the ordinate the probability of seeing a curve is plotted for five conditions (from left to right). In the first two conditions a curved trajectory is shown consisting of seven resp. five frames (c7, c5). In the third condition, three frames are shown (s3) and the last two conditions have a straight trajectory with five (s5) and seven frames (s7). The right side of Fig. 10 shows the reaction times for these two groups, the abscissa as above. The data are pooled over all SOAs, since this parameter did not influence the perceived path.

Discussion. There seem to be two groups showing different strategies, one group with an almost constant perception (six subjects) and another group (with only one subject, IK) with an interesting dependent behavior. The reports about the seen movements by subject IK is what we had expected: If all frames can be attributed to a straight path then subjects perceive just the straight path. This is in accordance to the phenomenon called 'visual momentum', which states that smooth trajectories are perceived preferably. But, adding frames so that the point moves on a curve, subjects should tend to report also a curved path and this probability should be directly connected to the length of the history. Surprisingly, only one subject shows this behaviour. Most of them nearly always report of having seen a straight movement. However, when

looking at the reaction times even this group seems to be effected by the kind of motion history, since there is a performance decrease of approximately 400 msec (condition c7 and c5 vs. condition s3, s5 and s7).

The RT differences for the second group are more pronounced (most notably in condition 2), but the impairment in the first group can perhaps be explained by subthreshold ambiguity. Due to these remarkable individual differences and the dichotomic distribution more subjects have to be recruited. Until now there is some evidence that the perceptual switching depends on the spatio-temporal memory span. Maybe only the last few hundred msec determine the perceived movement and therefore further experiments with smaller SOA are planned.

2.5 Spatio-Temporal Prototypes

This experiment addressed the issue of motion prototypes presenting very complex motion trajectories. The starting point was a pilot study (described in detail in [9]) in which subjects had to reproduce a complex motion shape on a touch screen.

The stimuli and results in Fig. 11 show that accuracy is mainly determined by features of the motion paths and thus point toward the existence of prototypical shapes like loops and kinks used by the visual system to classify a trajectory or a part of it. This encouraged us to start an experiment with stimuli that are simple enough to allow for quantitative results and that are complex enough so that conclusions about prototypes can still be drawn. Especially suited are the classes of stimuli which can be described as sequences of qualitative motion vectors (QMV) [9]. Several examples of these restricted trajectories are shown in Fig. 12. They consist of 16 segments of constant length and are used as stimuli.

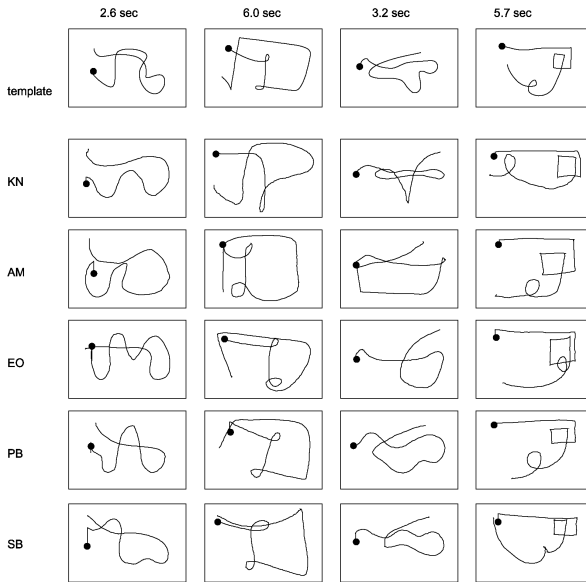


Fig. 11. Reproduction task: presented paths (templates) shown in the top row (with stimulus duration in seconds), reproduced paths below (results of five subjects).

We presented subsequently two trajectories as a first pair and after a short pause another two as a second pair (see Fig. 13, left). The subjects' task was to indicate whether there was a difference in the dot movements presented as first pair or in those of the second one. There were two classes of stimuli, some containing an element assumed to be a prototypical motion shape (the circular motion, stimulus class 1) and others lacking this property (stimulus class 2 in Fig. 12). The reference stimuli (duration 6.5 sec, constant speed of 4 deg/sec) had to be compared with three modified versions (test stimuli in Fig. 12). For them we varied the third (test 1), seventh (test 2) or 14th segment (test 3) about 45 (test 1) or 90 deg (test 2 and test 3).

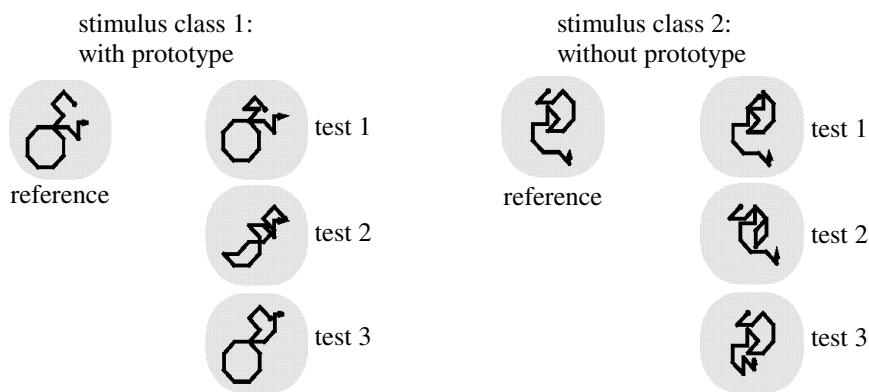


Fig. 12. Trajectories of QMV-stimuli.

Results. The percent correct scores for both stimuli classes can be seen in Fig. 13. There were 48 measurements in each conditions, since responses are pooled over subjects and repetitions.

Discussion. The main result is that the discrimination performance critically depends on the existence of motion prototypes. This is most obvious if we compare the performance in second condition, where subjects had to compare the reference with the second test stimulus. For the test stimulus of class 1 this means that the prototypical element is destroyed and the resulting difference can then be easily recognized. However, for the other stimulus (class 2) without such an element, the same amount of change results in a very poor discrimination performance. For this class of stimuli only the third variation leads to a comparably good performance (test 3 in Fig. 13). This can be explained because here the modified stimulus contains a zigzag-movement and this serves as a prototypical element.

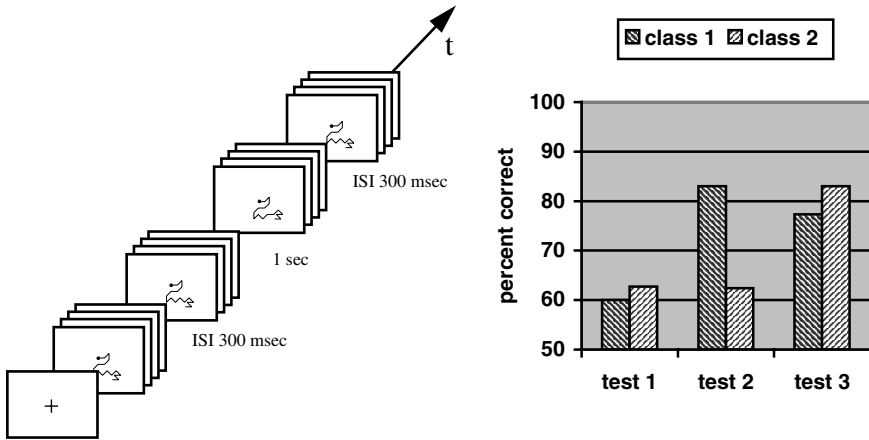


Fig. 13. Time course of the 2AFC experiment and results for four subjects.

In order to reveal more elements of this ‘shape vocabulary’ currently experiments with a number of supposed prototypes have been investigated. This experimental framework will also be extended to address explicitly the processing on different time scales and on several spatial granularities, e.g. by systematically varying moving time, number of segments, and allowed orientations.

3 A Hierarchical Neural Network Model

Besides the question of the features suitable for a shape vocabulary a further key question is how a sequence of elementary spatio-temporal features is processed by the visual system in order to get more complex trajectories on longer time scales like the characteristic walk of a person, the figural sequence of a conductor communicating with his orchestra, or a diving sequence with numerous turns and loops.

An analysis of the experimental results for an early memory stage has revealed inconsistencies arising with the dominant views of how information is represented and stored on this stage. Introducing a memory structure, which provides basic requirements for the processing of spatio-temporal information resolves these inconsistencies [16]. The key feature of this memory structure is the provision of an orthogonal access structure, which is achieved by mapping external time into an internal representation and temporal structure into locally distributed activities. This mapping enables a parallel access to a whole sequence of recently past visual inputs by higher level processing stages, which we call an orthogonal access memory (OAM).

In order to understand more about the structural constraints for the processing of higher-order sequences we started to analyze neural network models suitable for the representation, categorization and prediction of sequences on different time scales. There are several artificial neural network models described in the literature, which are capable of storing information for a certain period of time, e.g. *back-propagation through time* [17], *recursive cascading memory networks* and *time delay neural networks* [7]. They were often linguistically inspired and are used for speech

recognition. All these models use a supervised learning rule. Since we are interested in an architecture, which is able to work in an unsupervised manner, these models are not suited for our purposes. An unsupervised architecture that realizes different scales is, for instance, dynamic Self Organizing Maps (SOMs) as described in [12]. However, this architecture is not capable of predicting sequences. So we chose *Learning And Processing Sequences* (LAPS) [2] as a basis for our model. LAPS is a neural network architecture based on Kohonens SOMs [6] which are organized in a hierarchical manner with an additional feed-forward level connected to the system's output.

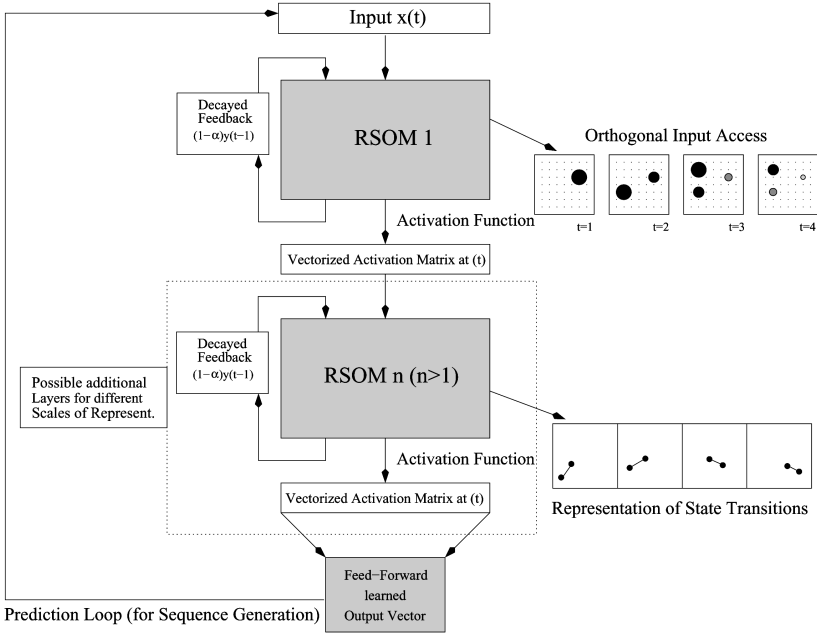


Fig. 14. The neural network model

We extended the original model in a way that it is capable of representing temporal and spatial knowledge on multiple scales (see Fig. 14). The result is a multi-layered structure in which each SOM represents a scale of abstraction regarding time and space [1]. Spatio-temporal properties are learned through feeding the output information of each SOM layer back to the next lower-order map. The model was further extended by an additional recurrent connection on the first SOM layer. This RSOM is capable of representing input sequence information through a decayed feedback of the activation of each neuron to its input.

The activation pattern of the first map represents the current event and the sequence of past events by a decayed feedback of the activation denoted by $y^{(1)}(t)$:

$$y_{ij}^{(1)}(t) = (1 - \alpha)y_{ij}^{(1)}(t-1) + \alpha \sum_k (x_k^{(1)}(t) - w_{i,j,k}^{(1)}(t)) \quad (1)$$

The parameter alpha models the behaviour of a single neuron with respect to time in that it delimits the interval in which an orthogonal access is possible and its value is

determined by experimental results (see above). The recurrent SOM enables an orthogonal access on a whole sequence of past elements or events as required by our suggested memory model (OAM) for an early representational stage. The sequence elements are ordered by their activation. After selecting the best matching unit with respect to the current input stimulus the decayed information from the past states are added. The development of the activation is illustrated in the following sketch (Fig. 15).

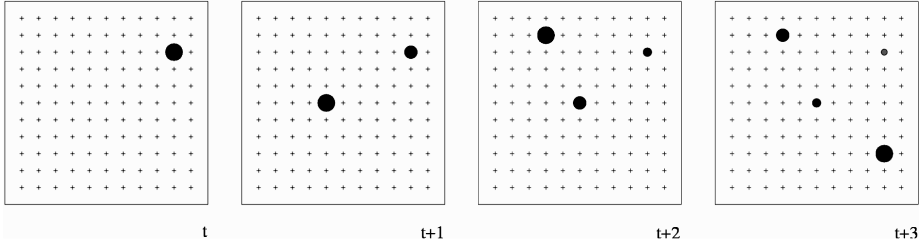


Fig. 15. Orthogonal access to a sequence of past events provided by the first layer: activation pattern of 11x11 neurons over time.

The second layer stores state transitions and further layers compute the higher-order relations between state transitions. The information, which is propagated from one layer to the next consists of the vectorised activation matrix of the Kohonen map on layer D concatenated with the output vector of layer $D+1$ at time $(t-1)$.

The vectorisation of the matrix $Y^{(D)}$ at time (t) is:

$$\text{lin}(Y^{(D)})(t) = y_{i+j \cdot n^{(D)}}^{(D)}(t) \quad (2)$$

with $D \in \{1, \dots, m\}$, and $n^{(D)}$ denoting the dimension of the systems input vector. The formal description of the input vector $i^{(d)}$ of SOM d , $d \in \{2, \dots, m\}$ is given by:

$$i_k^{(d)}(t+d) = y_{i+j \cdot n^{(D)}}^{(d-1)}(t+d) \oplus y_{i+j \cdot n^{(D)}}^{(d)}(t+d-1) \quad (3)$$

The activation of the subsequent layers is calculated like in standard SOM:

$$y_{ij}^{(d)}(t) = \sum_{k=1}^n (x_k^{(d)}(t) - w_{i,j,k}^{(d)}) \quad (4)$$

The best matching unit is selected by minimum search on the activation. To conserve sequence information for prediction it is necessary to learn the association between the systems output vector and the corresponding input vector. This is done in an extra layer by hebbian learning. The hebbian-learned output vector is necessary for prediction. By feeding this vector back to the input of the system, the system is capable of completing a sequence based on only a few starting values. The accuracy of the predicted sequence will decrease with increasing length.

4 Conclusion

The performance the human visual system achieves in processing and predicting spatio-temporal information is still superior to any technical systems (with exceptions of very restricted application ranges). This motivated us to analyze the visual systems ability for processing spatio-temporal patterns by psychophysical experiments and to transfer the gained insights to a hybrid modeling approach consisting of a neural network and a high-level qualitative stage. In our current psychophysical experiments we focused on two branches, namely, on the processing of motion shapes with differing complexity and, on the processing of motion information from different sensory systems. For the latter we found that if the consistency of auditory and visual information is consistent, i.e. if both sensory channels indicate the same motion direction, the resulting discrimination performance is better than in the inconsistent situation in which a visually presented object is approaching the subject and the auditory signal is moving away.

The psychophysical experiments in which we successively augmented the complexity of the motion shapes were started on a simple level of complexity where we compared the discrimination performance of changing direction of kinks vs. curves. Subsequently we conducted experiments in which we investigated the completion and prediction performance of partially occluded straight trajectories. One important result was the task dependent influence of spatial and temporal gaps. In a further experiment, based on an apparent motion paradigm, we analyzed the influence of the trajectory history on the extrapolation of more extended curved motion shapes. Finally, we investigated the recognition of partial changes in highly complex trajectories and identified some primitives and prototypical motion shapes. Our results allow first insights into the vocabulary of prototypical shapes used by the visual system and employ these as elements of the shape repertoire of our modeling approaches. For these approaches we considered two different methodologies: a propositional framework, suitable for qualitative spatio-temporal reasoning, and a neural network approach.

The neural network model is hierarchically structured and thus provides a basis for the representation and classification of spatio-temporal pattern on different layers of abstraction. It is based on multiple Kohonen SOM's and allows the unsupervised learning of spatio-temporal sequences. The properties of the first layer of this model provide an orthogonal storage structure, which enables the access to a whole sequence of past events. Previous work had shown that this access structure resolves inconsistencies arising with common views on internal representation and provides the necessary requirements for the processing of more intricate trajectory-like shapes. A feedback loop on the last level of the network allows us to predict sequences.

Our model will be developed further for technical applications such as robot navigation. In this case, the first computing level uses the input of motion selective filters corresponding to low level processing stages of the visual system. The next layer combines the output of these filters to pairs or longer sequences. In the highest processing level we have a representation of higher-order shapes. These sequences could be associated with the representation on a symbolic qualitative stage, which uses loops, corners or even higher abstractions of shapes, like complete motion plans. Thus, the system would have the ability to classify given spatio-temporal signals into higher abstractions of motor command sequences.

References

1. V. Baier, K. Schill, F. Röhrbein, W. Brauer. Processing of spatio-temporal structures: A hierarchical neural network model, *Dynamische Perzeption*, G. Baratoff, H. Neumann (eds.), 223-26, AKA, Akad.Verl.-Ges., Berlin, 2000.
2. G. Briscoe. Adaptive Behavioural Cognition, PhD thesis, Curtin University of Technology, School of Computing, Curtin, 1997.
3. Eisenkolb, A. Musto, K. Schill, D. Hernandez, W. Brauer. Representational levels for the perception of the courses of motion, *An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, Lecture Notes in Artificial Intelligence, C. Freksa, C. Habel, K. Wender (eds.), 1404, 129-55, Springer, Berlin, 1998.
4. A. Eisenkolb, K. Schill, F. Röhrbein, V. Baier, A. Musto, W. Brauer. Visual Processing and Representation of Spatio-Temporal Patterns. In C. Freksa, W. Brauer, C. Habel, K.F. Wender (eds.), *Spatial Cognition II - Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, 145-56, Berlin: Springer, 2000.
5. D. M. Green, J. A. Swets. *Signal Detection Theory and Psychophysics*, Wiley, 1966.
6. T. Kohonen. Spatio-temporal connectionist networks: A taxonomy and review, <http://hebb.cis.uoguelph.ca/~skremer/Teaching/27642/dynamic2/review.html>, 1998.
7. J. Miller. Channel interaction and the redundant-targets effect in bimodal divided attention. *J. Exp. Psych. HPP*, 17(1), 160-69, 1991.
8. A. Musto, K. Stein, A. Eisenkolb, T. Röfer, W. Brauer, K. Schill, From Motion Observation to Qualitative Motion Representation. In C. Freksa, W. Brauer, C. Habel, K.F. Wender (eds.), *Spatial Cognition II – Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, 115-26, Berlin: Springer, 2000.
9. Y. Nihei. A Preliminary Study on the Geometrical Illusion of Motion Path: The Kinetic Illusion. *Tohoku Psychologica Folia*, 32, 108-14, 1973.
10. C. Peterken, B. Brown, K. Bowman. Predicting the future position of a moving target. *Perception*, 20, 5-16, 1991.
11. C. M. Privitera, L. Shastri. Temporal compositional processing by a DSOM hierarchical model, Technical Reports 94704-1198, International Computer Science Institute Berkeley, Ca, 1996.
12. D. Raab. Statistical facilitation of simple reaction time. *Transact. N.Y. Acad. of Sci.* 43: 574-90, 1962.
13. V. S. Ramachandran, S. M. Anstis. Extrapolation of motion path in human visual perception, *Vision Research*, 23, 83-85, 1983.
14. F. Röhrbein, K. Schill, C. Zetsche. Intermodal Sensory Interactions for Ecologically Valid Intensity Changes as Caused by Moving Observers or Moving Objects. In H.H. Bülthoff, M. Fahle, K. Gegenfurthner, H. Mallot (eds.), *TWK 2000 – Beiträge zur 3. Tübinger Wahrnehmungskonferenz (62)*. Kirchentellinsfurth: Knirsch Verlag, 2000.
15. K. Schill, C. Zetsche. A model of visual spatio-temporal memory: the icon revisited, *Psychological Research*, 57, 88-102, 1995.
16. J. Tani. Model-based learning for mobil robot navigation from the dynamical systems perspective, *IEEE Trans. Systems, Man and Cybernetics (PartB)*, 26 (3), 421-36, 1996.
17. P. Wenderoth, M. Johnson. Relationship between the Kinetic, Alternating-Line, and Poggendorff Illusions: The Effects of Interstimulus Interval, Inducing Parallels, and Fixation. *Perception and Psychophysics*, 34 (3), 273-79, 1983.

Use of Reference Directions in Spatial Encoding

Constanze Vorwerk

Universität Bielefeld, SFB 360, Situated Artificial Communicators, Postfach 10 01 31,
33501 Bielefeld, Germany
Constanze.Vorwerk@uni-bielefeld.de

Abstract. Evidence is presented for the use of reference directions in verbal encoding and memory encoding. It is argued that reference directions (in linguistic spatial categorization as well as in memory encoding) are based on perceptually salient and distinguished orientations. A newly found spatial tilt effect for the sagittal in 3D visual space, that is reflected in different kinds of language processing, confirms a perceptual foundation of spatial language. It is proposed that direction is a qualitative attribute dimension, whose prototype values are not mean values or other characteristics of an empirical distribution but instead perceptually salient cognitive reference values. An account for angular bias effects in reporting location from memory is put forward and experimental results on the angular bias with and without physically present lines are presented.

1 Introduction

The main attributes used in specifying perceived location are *distance* and *direction*. A classification of spatial relations into distance relations and direction relations (sometimes called 'projective relations' since these are relations in terms of a particular perspective or point of view; Herskovits; 1986; Moore, 1976) holds for different areas of spatial cognition, including visual perception (Loomis, Da Silva, Philbeck & Fukusima, 1996), spatial knowledge about geographic regions (Thorndyke, 1981) or surroundings (Montello & Frank, 1996; Sadalla & Montello, 1989), cognitive maps (Anooshian & Siegel, 1985), verbally induced spatial images (Franklin & Tversky, 1990; Rinck, Hähnel, Bower & Glowalla, 1997), spatial memory (Huttenlocher, Hedges & Duncan, 1991), sensomotoric spaces (Paillard, 1987), formal spatial computational models (Gapp, 1995; Hernández, 1994) and spatial expressions of language (Landau & Jackendoff, 1993). Both distance and direction terms can be combined in verbal utterances (see Vorwerk & Rickheit, 2000).

Spatial location is coded as vector function. The spontaneous employment of a polar co-ordinate system has been shown for visual direction perception (Mapp & Ono, 1999), memory encoding of spatial location (Huttenlocher, Hedges & Duncan, 1991; Bryant & Subbiah, 1993) as well as verbal localization (Franklin, Henkel & Zangas, 1995; Gapp, 1995; Regier & Carlson, 2001; Vorwerk, 2001a). Therefore, direction can be regarded as angle (angular deviation) from a reference axis, whereas distance is determined as radial distance from the origin (see Fig. 1).

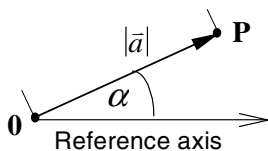


Fig. 1. Direction and distance: The position of a point **P** is given by directional angle α and magnitude $|\vec{a}|$ (= distance of **P** from the origin **0**) of the radius vector $\vec{OP}$.

This contribution looks into the *memory encoding of perceived direction* and its possible relation to *verbal localization* based on (direction) categorization processes. In order to be able to encode a perceived direction into memory or to categorize it linguistically, a certain direction must be chosen to serve as the axis or reference direction to which other directions can be referred to, i.e. deviation angles can be computed.

In principle, any clearly directed¹ axis of orientation can be used as a reference direction in relation to which other directions are judged; and there is extreme flexibility in specifying location w.r.t. reference direction in *perception*, *memory* and *language*. Different egocentric reference directions are available (see Müller, 1916; Pailard, 1991) as well as the axes of orientation of perceived or imagined objects (see Klatzky, 1998, e.g.). In addition, another (person's) perspective can be adopted (e.g., by means of mental rotation; see Herrmann & Graf, 1991; Huttenlocher & Presson, 1979; Shepard, 1988). Other examples include the direction of gravity (see Howard & Templeton, e.g.) or movement (Klein, 1979; Marcq, 1971).²

In *visual spatial perception*, certain (perceptually salient) directions are preferred as a reference direction, including the viewer's median plane, the line of vision (the direction in which the viewer is looking), and the eye-level horizontal, which is partly determined by the direction of the force of gravity (Matin, 1986). Other main reference values are provided by the apparent vertical (as determined by gravity and other influences) and ground level (Howard & Templeton, 1966). Perceptual reference directions seem to be directions "against which other visual directions or locations are reported with greatest acuity" (Matin, 1986, pp. 20/17).

In *verbal localization*, few linguistic expressions are available to cover an unlimited range of spatial locations. Therefore, a categorization process necessarily has to be involved in the naming of a perceived direction relation (as well as a distance relation). There is converging evidence that there are typicality gradients in the applicability of direction terms (see Vorwerg & Rickheit, 1998). Certain directions within a frame of reference are most easily and consistently encoded in language. With growing (angular) distance from them, typicality of perceived direction for a linguistic direction category gradually decreases, as exemplified by selection frequency,

¹ 'Directed' means 'directionally polarized' here.

² A systematic taxonomy of reference frames for direction is yielded by combining the criteria (1) determination of reference direction by either the observer/speaker, the addressee, or a third object, and (2) having the same or a different object as the reference object (a 'two-point' vs. a 'three-point localization'; Herrmann, 1990).

applicability rating and reaction time data. These findings, described in section 2, support the conclusion that in some frames of reference prototypical direction category values are perceptually salient orientations which are used in perceptual localization as well and which constitute representative cognitive reference directions to whom perceived direction relations may be referred.

In *spatial memory*, perceived locations may be encoded to provide a memory representation as a basis for comparison with a novel perceived location or for reproduction from memory. Reports from memory vary in precision and tendency. Certain directions within a frame of reference are encoded most accurately; and bias patterns can be described in relation to them. Not only do these directions correspond to perceptually salient orientations, but bias patterns observed also closely resemble known perceptual tilt effects. It will be argued that these perceptually distinguished orientations serve as cognitive reference directions in memory encoding, too. Evidence bearing on this issue is presented in section 3.

Finally, the findings presented will be discussed with regard to the question whether there are structural similarities between the memory representation of visual spatial relations and spatial language using a viewer-centered reference frame. Huttenlocher, Hedges, and Duncan (1991) and Crawford, Regier, and Huttenlocher (2000) contend that linguistic and non-linguistic categories do not correspond, but have an inverse relation such that the prototypes of linguistic spatial categories are boundaries in non-linguistic categories. Contrary to this conclusion, it will be argued that perceptually salient reference directions constitute basic frames of reference for linguistic encoding as well as memory encoding of perceived direction values.

2 Linguistic Encoding of Spatial Location

2.1 The Notion of a Frame of Reference in Spatial Cognition

Consideration of frames of reference is a key element in spatial cognition phenomena. The most basic notion of a spatial frame of reference is related to the fact that sameness of place or displacement and velocity can only be determined with respect to some other object or point in space. As Galilei (1632, transl. 1967, p. 116) put it: "Motion, in so far as it is and acts as motion, to that extent exists relatively to things that lack it; and among things which share equally in any motion, it does not act, and is as if it did not exist." This secondary object, with respect to which the spatial location of an *intended object* can be determined is often referred to as the '*reference object*'. Sometimes the term '*relatum*' is used to avoid the connotations of the word '*object*' as the *relatum* can also be, e.g., one's own body or another person.

With respect to a reference object, sameness of place can be judged, but not sameness of *direction* vs. *distance*. (Two locations may differ but have either the same direction or distance from a reference point.) Therefore, *directed orientations* must be used to define the orientation of main axes of the reference system. In a viewer-centered (or deictic) reference frame, these are determined by the perspective or point of view of the observer/speaker.

Both the relatum (or reference object) and the axes of orientation are required to be able to encode the direction of a perceived location. Relatum and axes of orientation are coordinated such that the origin of the (polar) reference frame is situated in the relatum (Vorwerg & Rickheit, 1999b; see also Gapp, 1995). In this work, ‘frame of reference’ denotes the position of the reference axes employed for specifying direction.

2.2 Reference Frames for Categorization

The notion of a frame of reference is also used in categorization and perception research. Following its usage in physics, gestalt psychology used the term to describe the fact that an entity in perception is qualified out of its relation to (preceding and concurrent) elements of the whole situation. Generally, perception can be understood as scaling w.r.t. a frame of reference (Thomas, Lusky, & Morrison, 1992).

Judgments of values on perceptual dimensions are often expressed in verbal, numerical, or other categories. Such a categorization is always based on a frame of reference: a set of values to which each given stimulus can be referred; e.g., focal colors, known size distribution of African elephants, or the loudness of a sound just heard. Standards for comparison may be given by memory representation and by the actual situation; and both work together in judgment of dimensional values. Although it is controversial to what extent the categorization of perceptual values is based on perceptual adaptivity itself relative to (semantic) category scale adjustment, it can be stated that both the perception and the categorization of dimensional degrees make use of reference frames.

The linguistic categorization of a perceived direction in space (which can be regarded as a dimensional attribute) necessitates a spatial as well as categorical frame of reference (see Vorwerg & Rickheit, 1998). That is, perceived direction values can be categorized according to their similarity (or angular proximity) to one of the main (half-)axes making up the spatial reference frame. For some frames of reference, namely the egocentric body-centered and the viewer-centered ones, perceptually salient orientations seem to provide the *cognitive reference directions* (Vorwerg & Rickheit, 1999b). These reference directions are proposed to act as ‘cognitive reference points’ (Rosch, 1975) forming the ‘prototype’ or ‘ideal type’ of a direction category akin to other not only quantitatively but also qualitatively variable perceptual dimensions, such as color or geometric forms (Vorwerg, 2001a).

In the following subsections, findings on the verbal localization within egocentric and viewer-centered reference frames are discussed with regard to the use of perceptually salient reference directions.

2.3 Egocentric Localization

In a study investigating the linguistic categorization (as well as memory encoding) of surrounding space, Franklin et al., 1995 asked subjects to describe the position of objects placed around themselves (in the horizontal plane) so that another person

could identify the place. Subjects were allowed to turn their heads and shoulders but not the rest of their bodies. Results showed the use of a trunk-centered frame of reference with the median plane (the midsagittal) as the reference direction defining the *front* half-axis of the frame (often referred to as *straight ahead*). In many of the verbal descriptions of location, 'front' was treated as a default value and not explicitly asserted, instead describing extent and direction of deviation from the front reference direction (e.g. *a little to the right*). The other horizontal reference directions (serving as prototypes for the categories 'back', 'left' and 'behind') can be derived from the front reference direction (within the horizontal plane). Linguistically, deviation from a reference direction or typicality for a spatial category can be expressed by hedge terms (such as *slightly*, *almost*, *a lot* or *directly*). Use of those hedges varied as a function of angular proximity to the nearest reference direction.

The median plane (with an egocentric midsagittal origin) is one of the main reference directions used for visual localization (Matin, 1986). Studies on pointing movements towards the (front) trunk midline on the basis of the mental representation of this line have shown that the midline is perceived as a straight line (Spidalieri & Sgolastra, 1997). It is concluded from the test results that somatosensory signals from the upper trunk and proprioceptive input from the neck contribute to the formation of the mental representation of the trunk midline. The trunk axis is actively stabilized during locomotion and it is used for target localization and movement trajectory planning as well as calculating leg position (see Massion, 1994, for a review). Besides that, it is an important factor (idiotropic vector) contributing to the perceived vertical (Luyat, Ohlmann & Barraud, 1997; Mittelstaedt, 1983; Neal, 1926). The trunk axis is coordinated with the head axis whose midsagittal is used as reference value for the vertical and for visual localization as well.

2.4 Deictic Localization

Experimental studies on the linguistic categorization of direction relations in the horizontal plane of 3D visual space have found that applicability of direction terms (such as *front*, *behind*, *left*, *right*) is a function of angular deviation from the reference directions determined by the line of vision (Vorweg, 2001a; Vorweg & Rickheit, 1999a; Vorweg, Socher, Fuhr, Sagerer & Rickheit, 1997).

In one series of experiments (Vorweg, 2001a), configurations of two objects (situated on a round table) were presented on an SGI computer screen using a pair of Crystal Eyes stereo glasses. Such a 3D display was necessary in order to be able to study the horizontal dimensions (FRONT-BEHIND and LEFT-RIGHT). Objects with no intrinsic front were chosen to avoid the use of conflicting intrinsic and deictic reference frames. The located object was a ring in order to prevent problems of axial alignment with the reference object. A (quadratic or rectangular) solid was used as a reference object in order to be able to investigate orientation effects (see Fig. 2).

In experiment 1, subjects *placed* an object with respect to a reference object as to fulfill a directional expression. Virtual reality technology has been applied for this investigation as it is difficult to study the manipulation of objects in 3D using a moni-

tor (even if the object configuration is perceived as three-dimensional). The positioning of the intended object turned out to be independent of the size of the reference object. The concentration of placements on the central axis originating from a coordination of direction of gaze and center of the reference object suggests the use of a main reference direction. This determination of the reference direction by the line of vision has been shown for 37 out of 39 subjects (the other 2 using a quasi-intrinsic frame assigning one of the relatum's sides the feature 'front').

In three categorization experiments, the reference object was located at the center of an array of 11 by 11 equidistant positions, one of which was occupied by the intended object in each trial of the experiments. Subjects either *named* the location of the intended object with relation to a given reference object by choosing from a given set of directional prepositions (experiment 2), or *rated* the applicability of a direction term for a location (experiment 3), or judged the category membership of a direction relation as quickly as possible (experiment 4). In order to be able to obtain *reaction time* data in this forth experiment, presentation on the SGI screen was linked with a reaction time measurement on a personal computer.

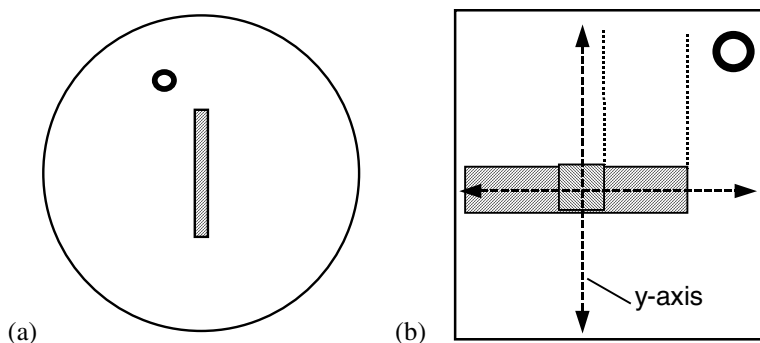


Fig. 2. (a) Schematic representation of a single object configuration as seen from above. (b) Schematic representation of the relation between proximity to central axis and proximity to proximal edge for two different reference objects with an intended object having the same position with regard to the central axis (from Vorwerg, 2001a).

Results showed that the *apparent sagittal* (similar to the apparent vertical in vertical space) determines the orientation of the main reference direction used. The primary factor defining the perceived sagittal seems to be the line of sight. Proportion of use and rated applicability of direction terms decreases as a function of deviation from a prototype value lying on the central axis (see Fig. 3). Frequency of choice and rating results are confirmed by reaction time data showing the significance of proximity to central axis for the processing of direction relations. One of the results obtained is a significant increase of reaction times with growing deviation from the central reference direction.

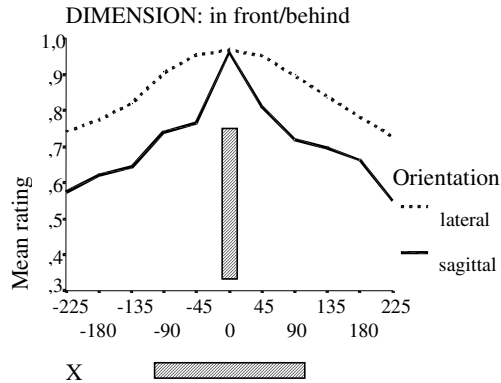


Fig. 3. Rated typicality of sagittal direction terms ('in front/behind') as a function of deviation from the y-axis (where $x=0$) for two different orientations of the reference object whose approximate extension in space (as seen from above) in relation to the intended object's positions on the x-axis is shown for reasons of comparison (from Vorwerg, 2001a).

These results correspond well with findings for visual localization: In three-dimensional visual space, one main reference direction is given by the (binocular) line of vision (see Matin, 1986; Müller, 1916). Whereas the vertical is determined by the force of gravity with the horizontal plane being derived from it, there is no universal, extrinsic reference direction analogous to gravity *within* the horizontal plane. Therefore, a local, perceptually salient directed orientation has to be found to determine the main horizontal reference direction. (Within the horizontal plane, the other three reference directions can be derived from the front reference direction.) One (egocentric) perceptually important reference direction is provided by the viewer's line of vision. Our data indicate that the speaker makes use of this reference direction in deictic localization. Direction categories are formed around perceptually salient prototype values, which seem to constitute frames of reference.

But, as the reference object usually is no single point in space, two important questions to ask are what influence the *extension* and the *orientation* of the relatum exert on the determination of the cognitive reference direction. The first question concerns the coordination between orientation of the reference direction as determined by line of sight and the orientation of the relatum. The origin of the reference frame could either lie on the central axis or on the proximal edge of the reference object (see Fig. 2; cf. Gapp, 1995; Regier & Carlson, 2001). Typicality ratings as well as relative frequency of direction-term choice turned out to be a function of both declination from the central axis and declination from the 'proximal axis' (see the difference between the two curves in Fig. 3). Both factors interact in the linguistic categorization of direction relations in deictic localization.

The second question considers the possible interaction of orientations. In addition to the cyclopean (binocular) line of vision providing the most important sagittal orientation in visual space, another orientation can be given by the longitudinal axis of a reference object of elongated shape (see Fig. 4).

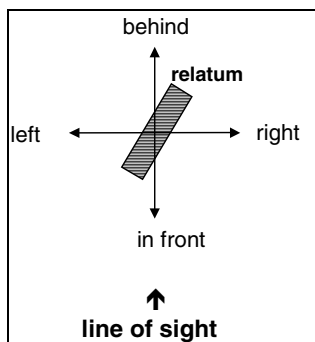


Fig. 4. The cyclopean line of sight provides the most important reference direction (the visual direction). Another reference orientation is the longitudinal axis of a reference object of elongated shape. Interactions of orientations may arise when both are neither collinear nor orthogonal.

Placements according to a given direction instruction ('put it behind/in front of X') have been shown to correspond to the best examples or prototypical values of a direction category (see also Logan & Sadler, 1996). Therefore the positions used with rotated reference objects (being neither collinear nor orthogonal to the line of sight) can be used to estimate orientation interactions. The results obtained w.r.t. this question in experiment 1 reveal a visual tilt effect of the reference object's orientation on the reference direction (Vorwerg, 2001a).

The deviation of the mean positions chosen by subjects follows the tilt of the *relatum* with small tilt angles and is opposite to the tilt of the *relatum* with large angles (see Fig. 5). That is, positions are displaced w.r.t. the true sagittal. With clockwise rotations of the *relatum*, this displacement is clockwise for small angles between line of sight and orientation of the *relatum* and counterclockwise for large angles. The same relation holds for counterclockwise rotations of the *relatum*. These tilt effects can be interpreted as a *tilt or rotation of the apparent sagittal* due to an interaction of the orientation of the line of vision and the orientation of the *relatum*. Similar tilt effects are known for the apparent vertical (in the fronto-parallel plane) depending on the tilt of a visual frame or background (Beh, Wenderoth & Purcell, 1971; Gibson, 1937; Witkin & Asch, 1948) or – in the opposite direction – the head's tilt (Aubert, 1861; Müller, 1916; see also Betts & Curthoys, 1998).

The observed visual tilt effect on the apparent sagittal was found again in the linguistic categorization of direction relation as measured by goodness rating, (forced-choice) frequency of use and reaction time (Vorwerg, 2001a). As the tilt effect can be interpreted as a perceptual effect, its effectiveness in different language tasks, such as the interpretation of a spatial expression or the use, applicability rating or verification of a spatial term for a given direction, can be regarded as strong evidence for a perceptual foundation of linguistic categorization of space and the use of perceptually based reference directions as prototypes or comparison values for linguistic categorization. Besides other factors influencing the choice of direction terms, including functional relations (Carlson-Radvansky, Covey & Lattanzi, 1999; Carlson-Radvansky &

Radvansky, 1996; Coventry, Carmichael & Garrod, 1994; Carmichael & Garrod, 1994) and verb semantics (e.g., Li, 1994), perceptual factors play an important role in linguistic localization.

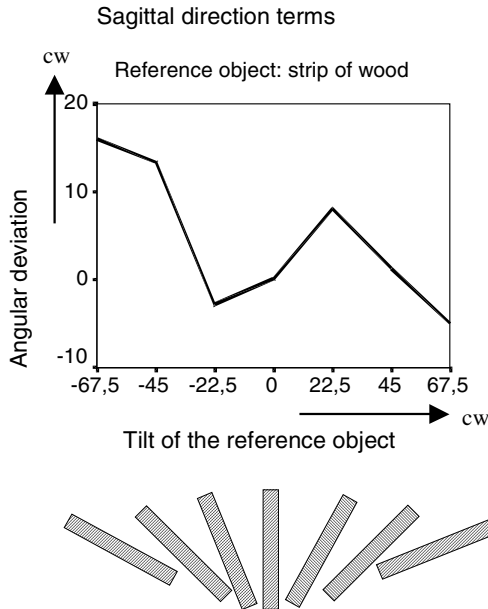


Fig. 5. Mean angular deviation of placements from the viewer-centered sagittal depending on the tilt of the reference object w.r.t. the sagittal (Vorwerg, 2001a). The approximate positions of the reference object on the horizontal plane are shown (as seen from above). Positive values indicate a clockwise tilt (of the reference object) and deviation (of placements) respectively.

2.5 Deictic Localization in 2D Space

In several studies, two-dimensional spatial configurations (consisting of a reference object and an intended object) have been used. In two-dimensional space, normally vertical and horizontal reference directions are used. In a viewer-centered (or deictic) frame of reference they are given by the vertical and the horizontal meridians of the visual field. When a reference object is used, its center often determines the origin of the co-ordinate system.

In the experiments on verbal localization in 2D visual space, either a reference object with an intrinsic vertical orientation (Crawford et al., 2000; Hayward & Tarr, 1995), a rectangular reference object without intrinsic top (Gapp, 1995; Regier & Carlson, 2001) or a small round reference object (Logan & Sadler, 1996; Zimmer, Speiser, Baus, Blocher & Stopp, 1998) was used. (In so far as reference objects with an axis of orientation were used, they were co-oriented with the subject's as well as a

possible frame's (the screen's, e.g.) orientation.) In all experiments, applicability of a direction term (such as *left*, *right*, *above*, *below*) turned out to be a function of angular deviation from the reference direction concerned (the vertical or the horizontal), measured by frequency of use (Hayward & Tarr; Zimmer et al.), goodness rating (Crawford et al.; Gapp; Hayward & Tarr; Logan & Sadler), and production latency (Zimmer et al.). It can be concluded again from these results that the axes of the spatial frame of reference form the prototypes for direction categorization.

The use of an external reference object necessitates the coordination of the (viewer-dependent) reference orientation and the position of the reference object. Therefore it has been investigated whether the *center* of the reference object or a *proximal* point of the reference object is used to define the (vertical or horizontal) reference direction. Based on work by Gapp (1995), whose data led him to conjecture that direction judgments are based on a proximal reference direction, Regier & Carlson (2001) varied angular deviation from centered vertical reference and from proximal orientation independently from each other. Their studies evidenced that direction term ratings are influenced by both deviation from proximal and from centered reference direction. This result for 2D visual space has been confirmed and extended for 3D visual space and the sagittal dimension by the experiments described in the previous subsection.

The use of the vertical and the horizontal as prototypical values for the linguistic categorization of spatial relations corresponds well with findings on the perceptual saliency of the vertical and the horizontal in vision. Horizontal and, especially, vertical orientations differ from other orientations within the fronto-parallel plane with regard to visibility (Ogilvie and Taylor, 1958, 1959), discriminability (Lashley, 1938), similarity judgments (Palmer, 1977) and symmetry perception (Lawson & Jolicoeur, 1998; Rock, 1973; Wenderoth, 1994). This correspondence between reference directions for linguistic categorization and perceptually salient orientations suggests that perceptually salient orientations are used as idealtypic values forming the prototypes of linguistic direction categories in deictic localization. By this account, other orientations or locations are perceived and categorized according to their declination from the reference directions which constitute the axes of the spatial reference frame.

3 Memory Encoding of Location

3.1 Use of Egocentric Co-ordinates

G. E. Müller (1916) investigated what egocentric co-ordinates are employed in learning and reproducing visually presented rows or complicated figures or in using diagrams or chromatisms. In the experiments conducted, he varied trunk, head and gaze position and as a result he found out that there are three frames of reference in relation to which a perceived object could be localized egocentrically.

Whereas the *head-centered* frame of reference changes its position in space with every movement of the head, the *trunk-centered* frame is defined by the trunk in its

normal posture according to the body schema (not its actual posture), the *view-centered* frame of reference can be defined as a co-ordinate system whose three axes are the line of sight, the vertical in the plane of view and a third straight line orthogonal to the other two. All three frames have been shown by Müller to be used for memory encoding and are known to be employed in perceptual localization (Matin, 1986).

3.2 Precision of Recalling Locations from Memory

In several studies, it has been shown that the accuracy with which subjects locate an object's former position is not uniform across different positions. For space around oneself in the horizontal plane, Franklin, Henkel & Zangas (1995) found absolute error to increase as a function of (angular) distance from the front pole, i.e. the trunk-defined midsagittal. This is a very important reference direction in perceptual localization and is often regarded as a kind of default value in verbal localization (see section 2.3).

For 2D visual space, Hayward & Tarr (1995) investigated memory representation by paradigms in which subjects either recalled the location of the intended object relative to another object or judged whether one of two objects presented sequentially was in the same position as the other or not. They observed that the region of greatest horizontal precision is directly vertical of the reference object and the region of greatest vertical precision is directly horizontal of it. These were the spatial positions where spatial terms had been judged to have high applicability. Moreover, these positions are located along the orientations that have been found to be perceptually salient and structuring visual space (see section 2.5; see also Vorweg & Rickheit, 1998).

3.3 Bias in Spatial Location Reproduction

In reports from memory, judgments of spatial location are frequently biased. Reproduced directions (of points) or orientations (of lines) are systematically misplaced away from or towards a reference value.

For surrounding space and a trunk-centered frame of reference, errors of reproduction have been found to be biased away from the front/back axis (Franklin, Henkel & Zangas, 1995). That is, errors were clockwise on the right side of front and counter-clockwise on the left side of front; and the opposite relation holds for the back region. This bias pattern can be described as a bias away from those reference directions constituting the primary horizontal axis, i.e. the sagittal axis.

For visual 2D space, in which a vertical and a horizontal dimension are differentiated, bias effects away from the cardinal axes have been observed. The cardinal axes can be given physically by two converging *vertical and horizontal reference lines* forming a right angle (). In these graph-like figures, the slopes of a third line converging on the origin of the axes (Schiano & Tversky, 1992; Tversky & Schiano, 1989) as well as positions of a single dot (Bryant & Subbiah, 1993) are remembered

systematically biased away from the cardinal axes. Moreover, a memory bias away from the *imaginary diagonal* (corresponding to the axis of symmetry of the angle made by the cardinal axes) has been demonstrated in some experiments (Bryant & Subbiah, 1993, Experiments 1 and 3; Schiano & Tversky, 1992, Experiments 1 and 2). Whether a repulsion effect from the imaginary diagonal occurs or not depends on the encoding strategy adopted by the viewer (cf. Schiano & Tversky, 1992; Tversky & Schiano, 1989).

The cardinal axes themselves can also be imaginary lines; i.e. they need not be drawn in the stimulus figures in order to produce a direction bias. Huttenlocher, Hedges and Duncan (1991) explored the reproduction of the location of a dot in a homogenous circle. They have found that subjects spontaneously impose horizontal and vertical axes through the center of the circle and misplace remembered dot locations away from each of these four half-axes. Within each quadrant there is a strong linear relation between angular error and actual angle. That is, the magnitude of bias is a function of the actual dot's angular deviation from axis with bias peaks near the axes. Dots that are located directly on the vertical and horizontal axes show little angular bias, especially on the vertical axis.

The use of the vertical and horizontal axes for memory encoding of location (as evidenced by the bias effects observed) corresponds well with both their importance in verbal localization and their special status in visual perception. But in order to be able to compare the use of reference directions - such as the visual vertical and horizontal - in memory and in language more detailed, the source of the direction bias has to be explored.

One of the most basic and most common frame of reference effects is the *contrast effect*. It is a context effect enhancing the difference between the subjective magnitude of a dimension value and a reference value. That is the perception and judgment of dimensional values is influenced by contrasting them with certain reference values. Also, sometimes *assimilation effects* (Steger, 1968) occur instead of contrast effects, a phenomenon that is not well understood. A contrast phenomenon bearing a close similarity to the reproduction bias described is *tilt contrast*: vertical test lines appear tilted away from a surrounding inducing grating for inducing angles up to about 60° from vertical (Smith & Wenderoth, 1999); lines converging and abutting to form an acute angle phenomenally repulse each other, especially if one line is either vertical or horizontal (Carpenter & Blakemore, 1973; Jastrow, 1893). Tilt contrast is generally reported to peak at small angles and has been shown for physically present as well as 'virtual' axes (e.g., Wenderoth, Johnston & van der Zwan, 1989).

The bias effects obtained for direction reproductions from memory can be described as *contrast effects* from the vertical or the horizontal, and in some conditions from the diagonals. And it can be concluded from several studies that perceptual mechanisms activated by encoding strategies contribute to the memory bias (cf. Bryant & Subbiah, 1993; Schiano & Tversky, 1992). However, a simple perceptual account in terms of a perceptual illusion would be faced not only with the problem of explaining the strategic effects but also with the question why the same illusion should not affect perception of the self-produced direction (such that a misplaced

reproduction of a location would be seen even further misplaced than the original location).

Huttenlocher, Hedges, and Duncan (1991) have argued that location is encoded hierarchically, i.e. at two levels – a category and a fine-grain level. Their model posits that encoding of fine-grain location is imprecise but unbiased with imprecision increasing by loss from memory, especially due to interference tasks between encoding and reproduction. Assuming that spatial regions (e.g., the quadrants of a circle into which it is divided by imposing horizontal and vertical axes) correspond to categories and that categories are represented by central values, the model posits that inexact fine-grain representation is combined with category level information. That is, reported location is supposed to be a kind of blending between an actual stimulus value and a category value, weighting them according to their associated inexactness. Based on this model of category effects on reports from memory, the authors come to the conclusion that *central values* within each quadrant of the circle form the prototypes for direction categories.

One implication of the model proposed by Huttenlocher et al. (1991) is that the prototypes of direction categories should lie along the obliques. This assumption is diametrically opposed to the findings concerning the linguistic categorization of 2D visual space (see section 2.5). Therefore, Crawford, Regier & Huttenlocher (2000) suggest that *linguistic and non-linguistic direction categories* do not correspond, but have an *inverse relation* such that the prototypes of linguistic categories are boundaries in non-linguistic categories.

In view of the compelling evidence indicating the use of the vertical and the horizontal as *perceptually salient reference directions* in perception as well as in language (cf. sections 1 and 2.5; for a review see Vorweg & Rickheit, 1998) it seems somewhat surprising that the direction prototypes for memory representations should lie along the obliques. An indeed, the accuracy with which locations on the vertical and the horizontal can be remembered has been interpreted as evidence for the idea that these directions serve as prototypes in memory encoding (cf. section 3.2). Huttenlocher et al. (1991) regard their bias data described above as supportive of their theoretically developed model since locations are systematically placed further towards the diagonal than the original dots. It is concluded from this that the diagonals serve as prototypes whose values are combined with the (imprecise remembered) actual values.

But one can just as well describe the bias found as systematical misplacement *away from the prototypic reference values* (with angular distance to them being a strong predictor of bias - being strongest near them). Therefore, an alternative account might consider the horizontal and vertical half-axes as prototypes (instead of boundaries) and *deviations* from prototypes could be encoded resulting in a cognitive enhancement of these deviations by *contrast* (similar to a schema-plus-tag-model). This account would attribute the obtained bias effects rather to encoding than to retrieval processes, contrary to the model proposed by Huttenlocher et al. (1991). It seems a more parsimonious assumption that the same kind of reference values is used in both memory and language encoding of visually perceived locations. Altogether, this assumption also fits into a coherent picture of visually based spatial cognition

including data on different frames of reference (e.g., the egocentric one) and categorization principles (which can be assumed to differ for quantitatively and qualitatively varying perceptual dimensions; see Vorwerk, 2001). Some experiments, presented in the following sections, were designed in order to explore in more detail the mechanisms involved in the memory encoding of direction relations (Vorwerk, 2003).

3.4 Bias Effects of Reference Lines

The systematic bias in memory for the location of a dot in a circle found by Huttenlocher et al. (1991) can be described as a contrast effect away from the vertical and the horizontal. The vertical and the horizontal might be used as *region boundaries* (Huttenlocher et al., 1991) or as *cognitive reference directions* (see also Vorwerk & Rickheit, 1999b). Some suggestions concerning the use of the horizontal or the vertical in location reproduction from memory may be gained from a direct comparison with dot placements from memory when either vertical and horizontal axes or diagonal axes are drawn (see Fig. 6). An experimental study addressed this question.

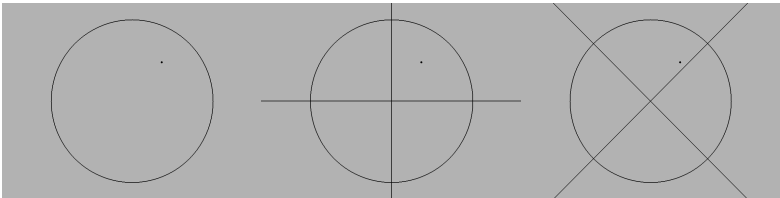


Fig. 6. Examples for stimuli in the three conditions.

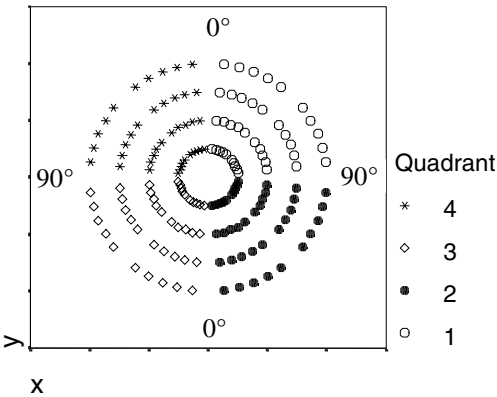


Fig. 7. Overview over the positions tested. Within each quadrant, 10 angular directions (spaced 7.5° apart with exception of values on the vertical or horizontal or on the diagonals) and 4 distance values were used.

The three conditions (no lines vs. straight lines vs. oblique lines) were varied between subjects. (For the sake of brevity, the vertical and the horizontal are referred to as 'straight lines' with 'straight' meaning 'level or upright' here.) In each condition, 160 positions were used (see Fig. 7). These were yielded by a combination of 4 (equidistant) distances and 40 angular directions. Angular directions were spaced 7.5° , but positions directly at one of the axes were left out. The stimuli were presented on a computer screen. In each trial, a circle and a dot within it were presented. The stimulus was presented for 600 ms followed by a visual mask image for 600 ms to ensure that subjects could not fixate on the position of the dot on the screen. Then the circle with or without lines (depending on condition) reappeared and the subject marked the location of the to-be-remembered dot with a mouse pointer.

The placement of dots from memory showed a bias in all conditions with bias being a function of angular deviation from the axes and reference lines. There is a bias away from the vertical and the horizontal in all three conditions and an additional bias away from the diagonals in condition 2 and 3 (the conditions with either vertical/horizontal or diagonal reference lines). Such a use of virtual lines has been shown for symmetry perception (Wenderoth, 1983) and for dot localization between a vertical and a horizontal line forming a right angle (Bryant & Subbiah, 1993). Another general effect is a reduction of bias (i.e. a flattening of the curve) near those reference directions (either vertical/horizontal or diagonal) that are not physically present.

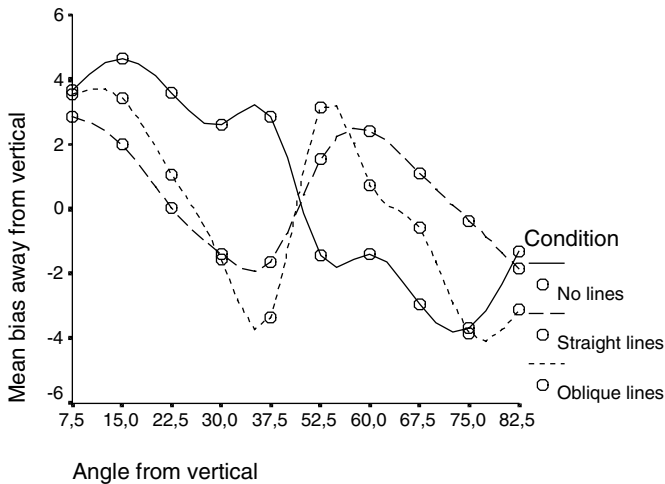


Fig. 8. Mean angular bias away from the vertical for data pooled together from all four quadrants (see Fig. 7). A positive value indicates a bias away from the vertical; a negative bias is towards the vertical (or away from the horizontal). Reference lines were used during encoding and retrieval.

Results show that all available reference directions are used to determine a perceived location. For a category-weighting model, these results would mean that the circle were divided into eight regions in case of reference lines. For a reference-directions

model, one might conclude that angular deviation from both neighboring reference direction is determined with acute angles being exaggerated during encoding into memory. For an encoding relative to reference directions, one can assume greater contrast effects from the vertical than from the horizontal. And that is indeed what we find for four out of five angles in the straight-lines condition and for two out of five angles in the no-lines condition (one more difference being tendentially significant). In order to be able to demonstrate potential verticality effects as such just described, bias is defined as bias away from the vertical, here (see Fig. 7 and Fig. 8). Thus a positive bias can be either clockwise or counterclockwise depending on whether the actual location of the dot is clockwise or counterclockwise deviating from the vertical. In other words, a positive bias is defined as angularly away from the vertical (and toward the horizontal).

On the assumption that reference directions are used in the encoding of location into memory, the present data support the conclusion that cognitive ('imaginary') reference directions are applied in a similar way as physically present reference lines or axes of symmetry between present reference lines. The deviation of an encoded location from a reference direction is exaggerated during encoding. The exaggeration is a function of angular deviation from a reference direction declining with greater angle. At positions very near a reference direction, a kind of attraction effect occurs assigning the position to the reference direction. The vertical has a special status in visual perception even compared to the horizontal and tends to cause greater contrast effects.

3.5 Encoding Effects on the Angular Bias

In another experiment, vertical and horizontal ('straight') lines vs. diagonal ('oblique') lines were presented only during the encoding of the dot's location. After the interference picture, just the circle reappeared in all conditions and the subject marked the location of the to-be-remembered dot with a mouse pointer. During retrieval, the lines were left out in order to investigate the role of encoding processes in the direction bias. Apart from that, the material and the procedure applied were the same as in experiment 1 (see section 3.4).

The results of this encoding-only experiment are quite similar to the findings for the experiment with reference lines during encoding and retrieval presented in section 3.4 (see Fig. 9; cf. Fig. 8). And there is a verticality effect for one out of five angles in the straight-lines condition (and one more tendency). Furthermore, the diagonal contrast effect is reduced for the 'straight lines' condition.

The reduction of bias (flattening of the curve) near those reference directions that are not physically present is again found for this experiment, similar to the first experiment with reference lines present both at encoding and at retrieval.

Contrary to what would be expected on the basis of a prototype-weighting model, 22.5° and 67.5° biases differ between both line conditions.

Results show that the direction bias depends on encoding conditions. The general form of the (primarily cubic) curves is not affected by retrieval conditions. As the

determination of all possible reference directions during retrieval is similarly difficult for all three conditions, differences between conditions can only be accounted for by perceptual encoding conditions. People seem to use as many perceptually salient reference directions as are available.

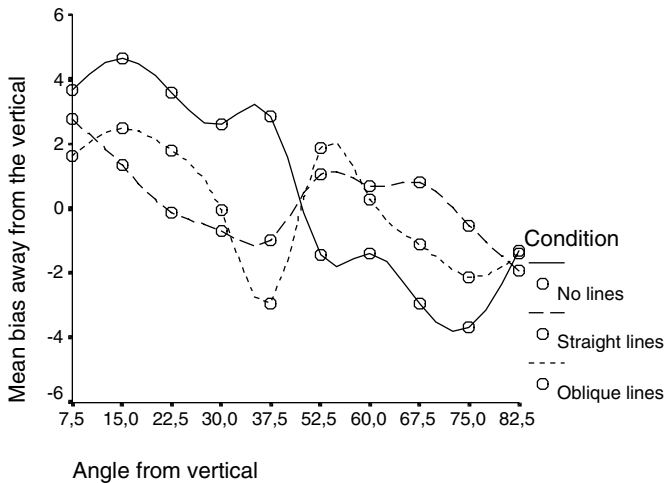


Fig. 9. Mean angular bias away from the vertical for data pooled together from all four quadrants (see Fig. 7). A positive value indicates a bias away from the vertical; a negative bias is towards the vertical (or away from the horizontal). Reference lines were used only during encoding. For comparison, the condition without lines is presented again.

3.6 Time Interval between Encoding and Retrieval

In a third experiment, a distractor time of 2000 ms between encoding and retrieval was used (instead of 2000 ms in the other experiments). After longer time intervals, it should be more difficult to reconstruct and use reference lines present during encoding but not during retrieval. By contrast, the use of cognitive reference directions (the vertical and the horizontal here) should be possible in a similar way as with brief intervals. Verticality effects (differences between bias from vertical and bias from horizontal) are expected to be more distinct with longer intervals, because deviations from the vertical should be given greater weight than deviations from the horizontal with more difficult reproduction conditions.

The enhancement of memory demands by lengthening the distractor time seems to reduce the importance of physically present reference lines in encoding (as long as these are not present at retrieval time) and to increase the use of the main reference directions (the vertical and the horizontal) instead (see Fig. 10).

Bias effects are reduced for the diagonals, especially in the 'oblique lines' condition. All angles show marked verticality effects for the two conditions 'without lines' and 'with straight lines', that is bias from vertical is stronger than from horizontal.

Again, 22.5° and 67.5° biases differ between both line conditions. On the basis of a prototype-weighting model they would be expected to correspond and to be approximately zero. If vertical/horizontal as well as diagonal reference directions were used as boundaries of spatial categories, the supposed central prototypes should be located at approximately 22.5° and 67.5° from the vertical. Therefore, it seems difficult to explain differences in the extent of bias between 'straight' and 'oblique' lines and between vertical and horizontal as a misclassification of some stimuli into a wrong category (cf. Huttenlocher et al., 1991). 22.5° and 67.5° angles do not seem to provide central values of categories.

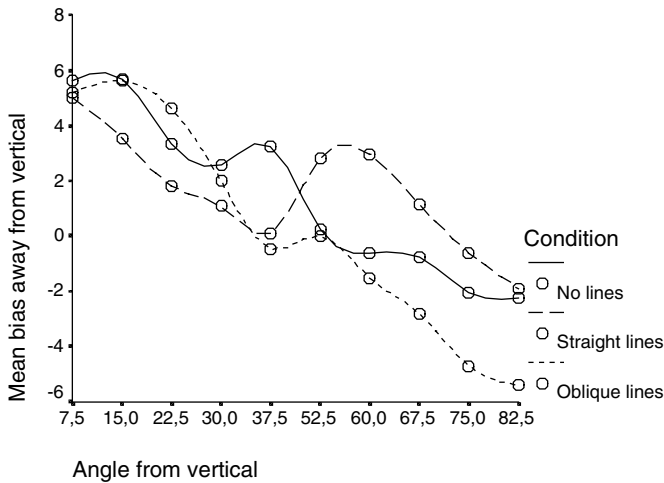


Fig. 10. Mean angular bias away from the vertical for data pooled together from all four quadrants (see Fig. 7). A positive value indicates a bias away from the vertical; a negative bias is towards the vertical (or away from the horizontal). Reference lines were used only during encoding. Time between encoding and retrieval was 2000 ms (compared to 600 ms in the other experiments).

Nevertheless, the general bias pattern corresponds to those found in other experiments. Data show contrast effects from both reference lines and virtual reference directions, a flattening of curves near virtual reference orientations and greater bias from vertical than from horizontal. The results are consistent with the view that physically present lines are only a short time available as reference directions, whereas cognitive reference directions, especially the vertical is given greater weight with longer intervals.

3.7 Imaginary Reference Lines

In a fourth experiment, subjects were asked to imagine either vertical and horizontal lines (condition 1) or diagonal lines (condition 2) within a circle presented on the

screen. After viewing the dot within the circle and a distractor picture, subjects first indicated the half-axis in proximity to which a dot was located and then marked the location of the to-be-remembered dot with a mouse pointer. Apart from that, the same procedure as in experiment 1 was applied (see section 3.4). The same positions as in the other experiments were studied (see Fig. 7).

The purpose of this instructional variation was to find out whether the specific imagination of reference lines although not present would affect the encoding of the stimuli. Results show strong contrast effects from the vertical and the horizontal for both conditions, but only small diagonal effects for the 'imaginary straight' condition and no diagonal effects for the 'imaginary oblique' condition (see Fig. 11). Furthermore, very strong verticality effects have been observed for both conditions (and for all angles investigated).

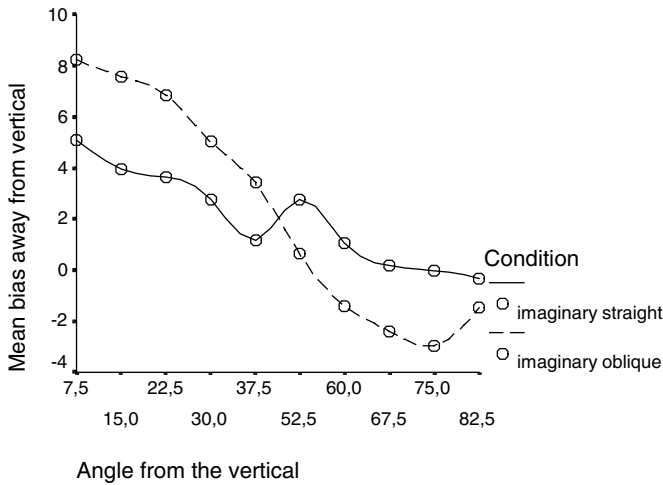


Fig. 11. Mean angular bias away from the vertical for data pooled together from all four quadrants (see Fig. 7). A positive value indicates a bias away from the vertical; a negative bias is towards the vertical (or away from the horizontal). No reference lines were used in this experiment. Subjects were asked to classify stimuli according to their proximity to either one of the vertical/horizontal half-axes or one of the diagonal half-axes.

A similar result as the one for the imaginary diagonal was obtained by Schiano and Tversky (1992) for the reproduction of a dot's location within a right-angle frame (see section 3.3) when subjects were instructed to use a diagonal-reference strategy (performing a diagonal comparison process). This finding is interpreted by the authors as reflecting an assimilation toward a cognitive reference value. However, a similar interpretation for the results presented here seems not adequate since the same instructional manipulation for the vertical and horizontal lines did not lead to a comparable effect. On the contrary, there is not only contrast from the horizontal and especially the vertical in both conditions, also the bias pattern for the 'imaginary straight' condition is similar to those observed for physically present lines. Moreover,

some subjects expressed their difficulty in using the diagonals as reference directions. Therefore, one possible account for the data is that subjects simply used the vertical and horizontal reference directions in encoding location in both conditions. (There was no explicit instruction to use one or the other in encoding, just to indicate the proximal half-axis). It is hypothesized that this lead to an additional interference task for the 'imaginary diagonal' condition causing a markedly greater bias than in all other experimental conditions. Additional support for this interpretation is gained from the fact that the vertical contrast effects outweigh the horizontal contrast effects. This verticality effect has been observed almost exclusively for the conditions without lines or with vertical and horizontal lines.

4 Discussion

A comparison of the findings for the linguistic categorization of direction relations and for the reproduction of direction values from memory supports the conclusion that cognitive reference directions are used in both verbal localization and memory encoding. Cognitive reference directions seem to be perceptually salient orientations. Several results on anisotropy in visual space confirm that there are preferred, salient and distinguished orientations within visual space, as exemplified by best resolution and detection acuity ('oblique effect'; Appelle, 1972), symmetry and form perception (e.g., Goldmeier (1937; Rock, 1973); the ability to adjust a line to the vertical or horizontal (as compared to other directions; e.g., Jastrow, 1893); comparison and discrimination of orientations (e.g. Attneave & Olson, 1967; Olson & Hildyard, 1977), and visual search (Treisman & Gormican, 1988). These salient orientations are the vertical and the horizontal for 2D visual space and especially the line of sight for 3D visual space.

In viewer-centered and egocentric frames of reference, cognitive reference directions are those in relation to which all actually perceived directions are judged and categorized according to their similarity to a reference direction. The angular proximity of a certain perceived direction to a cognitive reference direction affects the applicability of direction terms and the consistency, certainty and speed of linguistic categorization. These findings as well as the found (sagittal) visual tilt affecting different aspects of spatial language processing can be regarded as evidence for a perceptual foundation of spatial language. Phenomena demonstrating saliency of visual orientations have been observed for small children, non-human animals and basic visual processing stages. Therefore, the correspondence between perceptual saliency and use of reference directions for direction terms can not be attributed to linguistic categories affecting perception as discussed with respect to the Sapir-Whorf hypothesis for some other questions concerning the choice of direction terms.

The results presented show that perceptual factors play an important role in linguistic localization. Of course, there are other factor influencing the choice of direction terms, including functional relations (Carlson-Radvansky, Covey & Lattanzi, 1999; Carlson-Radvansky & Radvansky, 1996; Coventry, Carmichael & Garrod, 1994; Carmichael & Garrod, 1994) and verb semantics (cf. Li, 1994). Furthermore,

the impact of perceptual factors depends on the frame of reference used. The discussion here is concerned only with viewer-dependent and egocentric frames of reference. These are often regarded to be among the most basic reference frames. (Evidence for this assumption can be found, e.g., in developmental studies and also in etymology.) Nevertheless, other reference frames can be used as well and their choice depends on many factors, including cultural and maybe language factors. But the issue of choice of reference frame is beyond the scope of this contribution. The considerations with respect to the possible relation between spatial vision and spatial language are restricted to the question of categorising perceived direction relations *within one* certain (viewer-centered or egocentric) frame of reference.

Results for the encoding of location into memory suggest that the same perceptually salient orientations are used in memory encoding as well. One example is the finding that discrimination or reproductions from memory are more accurate for stimuli on the vertical or horizontal axis than for other locations (Hayward & Tarr, 1995). Furthermore, positions of cities on a map can be faster indicated for those located on the vertical or the horizontal axis thus providing evidence that also geographic knowledge is encoded relative to the main axes of a reference frame (Hintzman, O'Dell & Arndt, 1981). Alignment and rotation errors toward the vertical and the horizontal are evident in memory for location (Taylor, 1961; Tversky, 1981) and reproductions of tipped forms are often upright (Radner & Gibson, 1935).

Also the bias effects described in sections 3.3 to 3.7 can be regarded as evidence for the use of the vertical and the horizontal as important reference directions. They can easily be employed when no physically present orientations are available for comparison. Even though the exact mechanism underlying the bias effects is still under discussion and subject of experimental studies, different accounts seem to agree on the special status of the vertical and the horizontal in the encoding of (the directional aspects of) spatial location. And this might precisely be the reason why many results can be explained by different processing models. Whether the vertical and the horizontal are used as reference directions themselves, as hypothesized here, or define the boundaries of regions (Huttenlocher et al., 1991), both assumptions regard the vertical and the horizontal to be primary orientations, relative to which angular location is specified.

However, general considerations regarding model parsimony and consistency in accounting for related data might support the simpler assumption that one location is encoded only once in memory and that it is encoded in terms of angular deviation from a reference direction used to anchor orientation and direction perception. Additionally, some of the findings presented in sections 3.4 to 3.7 might possibly present difficulties to a category-weighting account, such as the vertical primacy found in different conditions using primarily the vertical and the horizontal as well as the differing bias effects for 22.5° locations in the two line conditions. Engebretson and Huttenlocher (1996) attributed a found greater bias for an imaginary vertical line (bisecting a right-angle) compared to an imaginary diagonal line (bisecting a differently oriented right-angle) to differing truncation processes due to different precision in imposing a vertical vs. a diagonal line. Such a factor can not account for the findings in section 3.6, because the bias pattern for the 'oblique lines' condition contains

almost no diagonal contrast effect. That would mean a complete loss of the assumed 'central prototype' values of 22.5° and 67.5° . For the other two conditions, the bias patterns are shifted away from the vertical.

One seeming contradiction, at first sight difficult to resolve, is the question why the employment of cognitive reference directions should cause assimilation effects in some cases and contrast effects in other cases. This problem might lead to the assumption that angular bias effects (if they are not perceptual tilt contrasts) have to be explained in terms of assimilation effects. However, the differential contrast vs. assimilation effects can be accounted for by fine-grained vs. coarse-grained encoding processes. In most cases, a coarse encoding of location will be sufficient and appropriate to the capacity of long-term memory. It can be hypothesized that assimilation effects occur for these coarse encoding processes facilitating the structuring of spatial representations. On the other hand, if attention is drawn especially to deviations from idealtypic, upright and straight configurations and orientations, this will lead to contrast effects by exaggerating deviations. This hypothesis is supported by results of Radner and Gibson (1935), who found that "forms objectively at an angle are often reproduced at upright but if the tip-character is noticed are usually reproduced with an exaggerated degree of tip". Indeed, they suggested that "with respect to orientation at least, a percept tends to occur at its perceptual center but that when a percept is experienced as departing from its center the eccentricity tends to increase" (p. 64). In the experiments conducted w.r.t. reproduction bias, attention is definitely drawn to angular deviation plus radial deviation as these are the only aspects by which items differ.

Generally, the categorization of angular direction is assumed to be based not on mean or central values depending on empirical distribution or dispersion of instances, but on proximity to *cognitive reference directions*. We have argued that angular location, which can be specified in terms of direction, is one of those attribute dimensions whose categories have an intrinsic qualitative distinctiveness and originate in and are constrained by perceptually salient stimuli (Vorwerg & Rickheit, 1998). In contrast to *qualitatively variable* attribute dimensions, such as direction, orientation or color, most attribute dimensions are *quantitatively variable*. This distinction has no relation to metric vs. categorical encoding (both kinds attribute dimensions can be encoded metrically or categorically); it simply draws on the fact that one direction or color value can not be said to be 'more' than another one, whereas a length or weight or brightness value can be 'more' or less than another 'one'. Qualitative dimensions seem to concern 'what kind' as opposed to 'how much' (see also Stevens, 1975, who used the terms *prothetic* vs. *metathetic continua*). Qualitatively and quantitatively variable attribute dimensions differ with regard to order of dimension values and categories, relation between magnitude (value) scale and category scale, semantic relation between category terms and reference values for categorization.

In quantitative dimensions, mean and range values of empirical distributions are of special importance in categorizing perceived values (see Vorwerg, 2001b). Reference values can be given by context or by memory. In qualitative dimensions however, reference values are provided by perceptually salient or cognitively distinguished ideal-typical values (see Wertheimer, 1912) or 'cognitive reference points' (Rosch,

1975). These values (e.g., focal colors, right angles, reference directions) are used as prototypes for categorization.

One fundamental distinction proposed for 'quantitatively' vs. 'qualitatively variable' attribute dimensions concerns the *principles of ratio forming*. Proportions between values are of decisive importance for achieving stability in perception. In many quantitative dimensions magnitude judgments follow a ratio scale (equal stimulus ratios produce equal subjective ratios; Stevens, 1975). Because of that, the magnitude ratio of two perceived values is independent of scale units. It can be judged more readily what weight feels twice as heavy as another weight, than what is their 'absolute' difference. Therefore the basis of the relation principle can be seen in invariant ratios. In a similar manner, scale-invariance has been proposed to be a unifying principle for the classical psychological laws (Chater & Brown, 1999). In contrast, data obtained by fractionation agree well with data obtained by equisection in qualitative dimensions (Stevens, 1975). Distance or intervals between two values of a qualitative dimensions can be judged successfully. Similarity can be determined as *distance similarity*. Comparison is based on intervals between values (contrary to proportions in quantitative dimensions). Given the fundamental importance of the relation principle in perception and categorization, it seems astonishing that ratio forming should not play a role in qualitative dimensions. I propose that qualitative dimension values in themselves are *based on a ratio of different subdimensions* (see Vorweg, 2001). These subdimensions can be provided by the three color primaries, the four taste primaries or the three spatial dimensions. The dimensions of angle, orientation, and direction rely on the ratio between two or more spatial dimension values. A value based on a ratio does not need a scale unit and can therefore be determined independent from other values of a dimension³.

Taken together, it can be concluded that the ratio principle is basic for both quantitative and qualitative dimensions and that quantitative and qualitative attribute dimension partly follow different categorization principles. Particularly important for the question of linguistic and memory encoding of direction (as compared to, e.g., distance) seems the use of different kinds of reference values for quantitative and qualitative dimensions.

References

- Anooshian, L. J. & Siegel, A. W. (1985). From cognitive to procedural mapping. In C. J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development: Progress in cognitive development research* (pp. 47-101). New York: Springer.
- Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation. The oblique 'effect' in man and animals. *Psychological Bulletin*, 78, 226-278.
- Attneave, F. & Olson, R. K. (1967). Discriminability of stimuli varying in physical and retinal orientation. *Journal of Experimental Psychology*, 74, 149-157.

³ In a similar way, the physical quantity 'angle' is 'dimensionless' because it is a quotient of two length quantities.

- Aubert, H. (1861). Eine scheinbare bedeutende Drehung von Objecten bei Neigung des Kopfes nach rechts oder links. *Virchows Archiv für pathologische Anatomie und Physiologie*, 20, 381-393.
- Beh, H., Wenderoth, P., & Purcell, A. (1971). The angular function of a rod-and-frame illusion. *Perception & Psychophysics*, 9, 353-355.
- Betts, G. A. & Curthoys, I. S. (1998). Visually perceived vertical and visually perceived horizontal are not orthogonal. *Vision Research*, 38, 1989-1999.
- Bryant, D. J. & Subbiah, I. (1993). Strategic and perceptual factors producing tilt contrast in dot localization. *Memory and Cognition*, 31, 773-784.
- Carlson-Radvansky, L. A., Covey, E. S. & Lattanzi, K. M. (1999). 'What' effects on 'where': Functional influences on spatial relations. *Psychological Science*, 10, 516-521.
- Carlson-Radvansky, L. A. & Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychological Science*, 7, 56-60.
- Carpenter, R. H. S. & Blakemore, C. (1973). Interactions between orientations in human vision. *Experimental Brain Research*, 18, 287-303.
- Chater, N. & Brown, G. D. A. (1999). Scale-invariance as a unifying psychological principle. *Cognition*, 69, B17-B24.
- Coventry, K. R., Carmichael, R. & Garrod, S. C. (1994). Spatial prepositions, object-specific function, and task requirements. *Journal of Semantics*, 11, 289-309.
- Crawford, L. E., Regier, T. & Huttenlocher, J. (2000). Linguistic and non-linguistic spatial categorization. *Cognition*, 75, 209-235.
- Engelbretson, P. H. & Huttenlocher, J. (1996). Bias in spatial location due to categorization: Comment on Tversky and Schiano. *Journal of Experimental Psychology: General*, 125, 96-108.
- Franklin, N., Henkel, L. A. & Zangas, T. (1995). Parsing surrounding space into regions. *Memory and Cognition*, 23, 397-407.
- Franklin, N. & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 63-76.
- Galilei, G. (1632). *Dialogue concerning the two chief world systems, Ptolemaic and Copernican*. Berkeley: University of California (Transl., 1967).
- Gapp, K. (1995). An empirically validated model for computing spatial relations. In I. Wachsmuth, C. Rollinger & W. Brauer (Eds.), *KI-95: Advances in Artificial Intelligence. Proceedings of the 19th Annual German Conference on Artificial Intelligence* (pp. 245-256). Berlin: Springer.
- Gibson, J. J. (1937). Adaptation, after-effect and contrast in the perception of tilted lines: II. Simultaneous contrast and areal restriction of the after-effect. *Journal of Experimental Psychology*, 20, 553-569.
- Goldmeier, E. (1937). Über Ähnlichkeit bei gesehenen Figuren. *Psychologische Forschung*, 21, 146-209.
- Hayward, W. G. & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition*, 55, 39-84.
- Hernandez, D. (1994). *Qualitative representation of spatial knowledge*. Berlin: Springer.
- Herrmann, T. (1990). Vor, hinter, rechts und links: das 6H-Modell. *Zeitschrift für Literaturwissenschaft und Linguistik*, 78, 117-140.
- Herrmann, T. & Graf, R. (1991). Ein dualer Rechts-Links-Effekt. *Zeitschrift für Psychologie, Suppl.* 11, 137-147.
- Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge: Cambridge University Press.

- Hintzman, D. L., O'Dell, C. S. & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology*, 13, 149-206.
- Howard, I. P. & Templeton, W. B. (1966). *Human spatial orientation*. New York: Wiley.
- Huttenlocher, J., Hedges, L. & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, 98, 352-376.
- Huttenlocher, J. & Presson, C. C. (1979). The coding and transformation of spatial information. *Cognitive Psychology*, 11, 375-394.
- Jastrow, J. (1893). On the judgment of angles and positions of lines. *The American Journal of Psychology (Reproduction 1966, ed. by G. S. Hall)*, 5, 214-248.
- Klatzky, R. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel & K. F. Wender (Eds.), *Spatial cognition. An interdisciplinary approach to representing and processing spatial knowledge* (pp. 1-17). Berlin: Springer.
- Landau, B. & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217-265.
- Lashley, K. S. (1938). The mechanism of vision: XV. Preliminary studies of the rats' capacity for detailed vision. *Journal of General Psychology*, 18, 123-193.
- Lawson, R. & Jolicoeur, P. (1998). The effects of plane rotation on the recognition of brief masked pictures of familiar objects. *Memory & Cognition*, 26, 791-803.
- Li, J. (1994). *Räumliche Relationen und Objektwissen am Beispiel 'an' und 'bei'*. Tübingen: Gunter Narr.
- Logan, G. D. & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett (Eds.), *Language and space* (pp. 493-529). Cambridge, MA: MIT Press.
- Loomis, J. M., Da Silva, J. A., Philbeck, J. W. & Fukusima, S. S. (1996). Visual perception of location and distance. *Current Directions in Psychological Science*, 3, 72-77.
- Luyat, M., Ohlmann, T. & Barraud, P.A. (1997). Subjective vertical and postural activity. *Acta Psychologica*, 95, 181-193.
- Mapp, A. P. & Ono, H. (1999). Wondering about the wandering cyclopean eye. *Vision Research*, 39, 2381-2386.
- Marcq, P. (1971). Structure d'un point particulier du système des prépositions spatiales en latin classique. *La Linguistique. Revue Internationale de Linguistique Générale*, 7, 81-92.
- Massion, J. (1994). Postural control system. *Current Opinion in Neurobiology*, 4, 877-887.
- Matin, L. (1986). Visual localization and eye movements. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 1: Sensory processes and perception* (pp. 20/1-20/45). New York: Wiley.
- Mittelstaedt, H. (1983). A new solution to the problem of verticality. *Naturwissenschaften*, 70, 272-281.
- Montello, D. R. & Frank, A. U. (1996). Modeling directional knowledge and reasoning in environmental space: Testing qualitative metrics. In J. Portugali (Ed.), *The construction of cognitive maps* (pp. 321-344). Dordrecht: Kluwer Academic Publishers.
- Moore, G. T. (1976). Theory and research on the development of environmental knowing. In G. T. Moore & R. G. Golledge (Eds.), *Environmental knowing* (pp. 138-164). Stroudsburg, Penn.: Dowden, Hutchinson & Ross.
- Müller, G. E. (1916). Über das Aubertsche Phänomen. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 49, 109-244.
- Neal, E. (1926). Visual localization of the vertical. *The American Journal of Psychology*, 37, 287-291.

- Ogilvie, J. C. & Taylor, M. M. (1958). Effects of orientation of the visibility of a fine line. *Journal of the Optical Society of America*, 48, 628-629.
- Ogilvie, J. C. & Taylor, M. M. (1959). Effect of length on the visibility of a fine line. *Journal of the Optical Society of America*, 49, 898-900.
- Olson, D. R. & Hildyard, A. (1977). The mental representation of oblique orientation. *Canadian Journal of Psychology*, 31, 3-13.
- Paillard, J. (1987). Cognitive versus sensorimotor encoding of spatial information. In P. Ellen & C. T. Blanc (Eds.), *Cognitive processes and spatial orientation in animal and man* (pp. 43-77). Dordrecht: Martinus Nijhoff Publishers.
- Paillard, J. (1991). Motor and representational framing of space. In Paillard, Jacques (Ed.), *Brain and space* (pp. 163-182). Oxford: Oxford University Press.
- Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9, 441-474.
- Radner, M. & Gibson, J. J. (1935). Orientation in visual perception. The perception of tip-character in forms. *Psychological Monographs*, 46, 48-65.
- Regier, T. & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, 130, 273-298.
- Rinck, M., Hähnel, A., Bower, G. H. & Glowalla, U. (1997). The metrics of spatial situation models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 622-637.
- Rock, I. (1973). *Orientation and form*. New York: Academic Press.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7, 532-547.
- Sadalla, E. K. & Montello, D. R. (1989). Remembering changes in direction. *Environment and Behavior*, 21, 346-363.
- Schiano, D. J. & Tversky, B. (1992). Structure and strategy in encoding simplified graphs. *Memory and Cognition*, 20, 12-20.
- Shepard, R. N. (1988). The role of transformations in spatial cognition. In J. Stiles-Davis, M. Kritchewsky & U. Bellugi (Eds.), *Spatial cognition. Brain bases and development* (pp. 81-110). Hillsdale, N.J.: Lawrence Erlbaum.
- Smith, S. & Wenderoth, P. (1999). Large repulsion, but not attraction, tilt illusions occur when stimulus parameters selectively favour either transient (M-like) oder sustained (P-like) mechanisms. *Vision Research*, 39, 4113-4121.
- Spidalieri, G. & Sgolastra, R. (1997). Psychophysical properties of the trunk midline. *Journal of Neurophysiology*, 78, 545-549.
- Steger, J. A. (1968). The reversal of simultaneous contrast. *Psychological Bulletin*, 70, 774-781.
- Stevens, S. S. (1975). *Psychophysics. Introduction to its perceptual, neural, and social prospects*. New York: John Wiley & Sons.
- Taylor, M. M. (1961). Effect of anchoring and distance perception on the reproduction of forms. *Perceptual and Motor Skills*, 12, 203-230.
- Thomas, D. R., Lusky, M. & Morrison, S. (1992). A comparison of generalization functions and frame of reference effects in different training paradigms. *Perception & Psychophysics*, 51, 529-540.
- Thorndyke, P. W. (1981). Distance estimations from cognitive maps. *Cognitive Psychology*, 13, 526-550.
- Treisman, A. M. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15-48.
- Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13, 407-433.

- Tversky, B. & Schiano, D. (1989). Perceptual and conceptual factors in distortions in memory graphs and maps. *Journal of Experimental Psychology: General*, 118, 387-398.
- Vorwerg, C. (2001a). *Raumrelationen in Wahrnehmung und Sprache. Kategorisierungsprozesse bei der Benennung visueller Richtungsrelationen*. Wiesbaden: Deutscher Universitätsverlag.
- Vorwerg, C. (2001b). Objektattribute: Bezugssysteme in Wahrnehmung und Sprache. In L. Sichelschmidt & H. Strohnner (Eds.), *Sprache, Sinn und Situation* (pp. 59-74). Wiesbaden: Deutscher Universitätsverlag.
- Vorwerg, C. (2003). *Contrast effects in the memory encoding of direction relations*. Manuscript in preparation.
- Vorwerg, C. & Rickheit, G. (1998). Typicality effects in the categorization of spatial relations. In C. Freksa, C. Habel & K. F. Wender (Eds.), *Spatial cognition. An interdisciplinary approach to representing and processing spatial knowledge* (pp. 203-222). Berlin: Springer.
- Vorwerg, C. & Rickheit, G. (1999a). Richtungsausdrücke und Heckenbildung beim sprachlichen Lokalisieren von Objekten im visuellen Raum. *Linguistische Berichte*, 178, 152-204.
- Vorwerg, C. & Rickheit, G. (1999b). Kognitive Bezugspunkte bei der Kategorisierung von Richtungsrelationen. In G. Rickheit (Ed.), *Richtungen im Raum* (pp. 129-165). Wiesbaden: Westdeutscher Verlag.
- Vorwerg, C. & Rickheit, G. (2000). Repräsentation und sprachliche Enkodierung räumlicher Relationen. In C. Habel & C. von Stutterheim (Eds.), *Räumliche Konzepte und sprachliche Strukturen* (pp. 9-44). Tübingen: Niemeyer.
- Vorwerg, C., Socher, G., Fuhr, T., Sagerer, G. & Rickheit, G. (1997). Projective relations for 3D space: Computational model, application, and psychological evaluation. *Proceedings of AAAI-97*. Cambridge, MA: AAAI Press/MIT Press, 159-164.
- Wenderoth, P. (1983). Identical stimuli are judged differently in the orientation and position domains. *Perception & Psychophysics*, 33, 399-402.
- Wenderoth, P. (1994). The salience of vertical symmetry. *Perception*, 23, 221-236.
- Wenderoth, P., Johnstone, S. & van der Zwan, J. (1989). Two-dimensional tilt illusions induced by orthogonal plaid patterns: Effects of plaid motion, orientation, spatial separation, and spatial frequency. *Perception*, 18, 25-38.
- Wertheimer, M. (1912). Über das Denken der Naturvölker. Zahlen und Zahlgebilde. *Zeitschrift für Psychologie*, 60, 321-378.
- Witkin, H. A. & Asch, S. E. (1948). Studies in space orientation. III. Perception of the upright in the absence of a visual field. *Journal of Experimental Psychology*, 38, 603-614.
- Zimmer, H. D., Speiser, H. R., Baus, J., Blocher, A. & Stopp, E. (1998). The use of locative expressions in dependence of the spatial relation between target and reference object in two-dimensional layouts. In C. Freksa, C. Habel & K. F. Wender (Eds.), *Spatial cognition. An interdisciplinary approach to representing and processing spatial knowledge* (pp. 223-240). Berlin: Springer.

Reasoning about Cyclic Space: Axiomatic and Computational Aspects

Philippe Balbiani¹, Jean-François Condotta², and Gérard Ligozat²

¹ Institut de recherche en informatique de Toulouse
118 route de Narbonne, 31062 Toulouse Cedex 4, France

² Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur
BP 133, 91403 Orsay Cedex, France

Abstract. In this paper we propose models of the axioms for linear and cyclic orders. First, we describe explicitly the relations between linear and cyclic models, from a logical point of view. The second part of the paper is concerned with qualitative constraints: we study the cyclic point algebra. This formalism is based on ternary relations which allow to express cyclic orientations. We give some results of complexity about the consistency problem in this formalism. The last part of the paper is devoted to conceptual spaces. The notion of a conceptual space is related to the complexity properties of temporal and spatial qualitative formalisms, including the cyclic point algebra.

1 Introduction

Much attention in the domain of qualitative temporal and spatial reasoning has been devoted to the study of spaces which are ultimately based on some version of a Euclidean space: Allen's calculus [1] is the qualitative study of pairs of points in the 1-D Euclidean space, the real line; the Cardinal Direction calculus [8,20], the n -point calculus [4], the rectangle calculus [3], the n -block calculus [5], the line segments calculus [22], refer to entities in (Cartesian products of the real line), which is an unbounded, dense linear ordering.

There are however good reasons for considering spaces which are not based on linear orderings. The set of directions around a reference points has a cyclic, rather than a linear structure. Schlieder's concepts of orientation and panoramas [25,26] are examples of proposals for reasoning about cyclic situations. Such is Röhrig's theory CycOrd [24] and work of Sogo *et al.* [28]. More recently, Cohn and Isli [12] have considered points on a circle and the ternary relations between them, obtaining substantial results about the complexity of the corresponding calculi. Finally, the binary relations between intervals on a circle have been considered [7]. If we think of the particular field of applications to reasoning about geographical or cartographic entities, it is clear that many applications may need to consider cycles such as parallels or meridians on the Earth's surface.

When studying spaces with a cyclic structure, it seems quite reasonable not to consider the cyclic case as a *tabula rasa*. After all, from a topological point of view, a circle is easily derivable from a segment (or a line) by identifying the end-points. Conversely, cutting a circle makes it into a line. This is the intuition behind the work presented in this paper: The idea is to exploit, as much as possible, the relationships between linear

and cyclic models. Technically, this will also involve the relationships between binary relations between points on a line, which are enough to characterize the “qualitative” relation between them, and ternary relations between three points on a circle, which are necessary for the analogous characterization.

The structure of the paper is as follows: Firstly, we describe explicitly the relations between linear and cyclic models, from a logical point of view. The main result is that, in the same way as there is basically one countable model of an unbounded, dense and linear ordering (Cantor’s theorem), a similar result obtains for suitable well chosen axioms involving one “betweenness” ternary relation between points on a circle (the main condition here is to have at least two points, plus density). Then, we consider six possible ternary relations between three points on a circle, which are jointly exhaustive and pairwise disjoint (JEPD) relations, and develop a qualitative calculus, and examine the problem of determining consistency for the corresponding constraint networks. We describe various subsets where we can prove either tractability or NP-completeness. Finally, in a last section, we consider the problem of extending the known complexity results for the linear calculi to the cyclic cases. Although the initial results (about Allen’s algebra) were first proved using logical tools, it appears that most of them can also be expressed in geometric and topological terms, which can be also understood as particular cases of conceptual spaces introduced by Gärdenfors. We present the basic notions of the framework of conceptual spaces, in relation to the characterization of tractable sub-classes. We then speculate on the possibility of extending the geometric and topological characterizations of tractable classes to the cyclic case.

2 Models of Axioms for Linear and Cyclic Orders

This section is devoted to the semantical analysis of the relationship between linear orders and cyclic orders.

2.1 Linear Orders

A linear order is a structure of the form $\mathcal{M} = (\mathcal{T}, <)$ where $\mathcal{T}$ is a set of points and $<$ is a binary relation on $\mathcal{T}$ subject to the following universal conditions for all $x, y, z \in \mathcal{T}$:

- Not $x < x$;
- If $x < y$ and $y < z$ then $x < z$;
- Either $x = y$ or $x < y$ or $y < x$.

A linear order $\mathcal{M} = (\mathcal{T}, <)$ is dense if it satisfies the following principle:

- For all $x, y \in \mathcal{T}$, if $x < y$ then there is $z \in \mathcal{T}$ such that $x < z$ and $z < y$.

A linear order $\mathcal{M} = (\mathcal{T}, <)$ is unbounded if it satisfies the following principles:

- For all $x \in \mathcal{T}$, there is $y \in \mathcal{T}$ such that $y < x$;
- For all $x \in \mathcal{T}$, there is $y \in \mathcal{T}$ such that $x < y$.

Before turning to cyclic orders and to a detailed investigation of their relationship with linear orders, let us remind the reader of the following result.

Proposition 1. *Let $\mathcal{M}_1 = (\mathcal{T}_1, <_1)$ and $\mathcal{M}_2 = (\mathcal{T}_2, <_2)$ be two countable linear orders. If $\mathcal{M}_1$ and $\mathcal{M}_2$ are dense and unbounded then they are isomorphic.*

Proof. By Cantor's well known zig-zag argument. $\square$

In technical terminology, the set of all dense and unbounded linear orders is countably categorical. Hence as far as countable structures are concerned, there is only one dense and unbounded linear order, the structure $(\mathbf{Q}, <)$ of the rational numbers.

2.2 Cyclic Orders

A cyclic order is a structure of the form $\mathcal{M} = (\mathcal{T}, \prec)$ where $\mathcal{T}$ is a nonempty set of points and $\prec$ is a ternary relation on $\mathcal{T}$ subject to the following universal conditions for all $x, y, z, t \in \mathcal{T}$:

- Not $\prec(x, y, y)$;
- If $\prec(x, y, z)$ and $\prec(x, z, t)$ then $\prec(x, y, t)$;
- If $x \not\prec y$ and $x \not\prec z$ then either $y = z$ or $\prec(x, y, z)$ or $\prec(x, z, y)$;
- $\prec(x, y, z)$ iff $\prec(y, z, x)$ iff $\prec(z, x, y)$.

A cyclic order $\mathcal{M} = (\mathcal{T}, \prec)$ is standard if it satisfies the following principles:

- For all $x, y \in \mathcal{T}$, if $x \not\prec y$ then there is $z \in \mathcal{T}$ such that $\prec(x, z, y)$;
- For all $x, y \in \mathcal{T}$, if $x \not\prec y$ then there is $z \in \mathcal{T}$ such that $\prec(x, y, z)$.

Referring to propositions 1, 4 and 6, we easily obtain a proof of the following result.

Proposition 2. *Let $\mathcal{M}_1 = (\mathcal{T}_1, \prec_1)$ and $\mathcal{M}_2 = (\mathcal{T}_2, \prec_2)$ be two countable cyclic orders. If $\mathcal{M}_1$ and $\mathcal{M}_2$ are standard then they are isomorphic.*

Proof. Assume that $\mathcal{M}_1 = (\mathcal{T}_1, \prec_1)$ and $\mathcal{M}_2 = (\mathcal{T}_2, \prec_2)$ are standard. Let ∞_1 be a point such that $\infty_1 \in \mathcal{T}_1$ and ∞_2 be a point such that $\infty_2 \in \mathcal{T}_2$. Let $\mathcal{M}'_1 = (\mathcal{T}'_1, <'_1)$ be the linear order on $\mathcal{M}_1 = (\mathcal{T}_1, \prec_1)$ and ∞_1 and $\mathcal{M}'_2 = (\mathcal{T}'_2, <'_2)$ be the linear order on $\mathcal{M}_2 = (\mathcal{T}_2, \prec_2)$ and ∞_2 . By proposition 4, $\mathcal{M}'_1 = (\mathcal{T}'_1, <'_1)$ and $\mathcal{M}'_2 = (\mathcal{T}'_2, <'_2)$ are dense and unbounded. By proposition 1, $\mathcal{M}'_1 = (\mathcal{T}'_1, <'_1)$ and $\mathcal{M}'_2 = (\mathcal{T}'_2, <'_2)$ are isomorphic. Let ∞'_1 be a point such that $\infty'_1 \in \mathcal{T}'_1$ and ∞'_2 be a point such that $\infty'_2 \in \mathcal{T}'_2$. Let $\mathcal{M}''_1 = (\mathcal{T}''_1, \prec''_1)$ be the cyclic order on $\mathcal{M}'_1 = (\mathcal{T}'_1, <'_1)$ and ∞'_1 and $\mathcal{M}''_2 = (\mathcal{T}''_2, \prec''_2)$ be the cyclic order on $\mathcal{M}'_2 = (\mathcal{T}'_2, <'_2)$ and ∞'_2 . We see without difficulty that $\mathcal{M}''_1 = (\mathcal{T}''_1, \prec''_1)$ and $\mathcal{M}''_2 = (\mathcal{T}''_2, \prec''_2)$ are isomorphic. By proposition 6, $\mathcal{M}_1 = (\mathcal{T}_1, \prec_1)$ and $\mathcal{M}''_1 = (\mathcal{T}''_1, \prec''_1)$ are isomorphic and $\mathcal{M}_2 = (\mathcal{T}_2, \prec_2)$ and $\mathcal{M}''_2 = (\mathcal{T}''_2, \prec''_2)$ are isomorphic. Hence $\mathcal{M}_1 = (\mathcal{T}_1, \prec_1)$ and $\mathcal{M}_2 = (\mathcal{T}_2, \prec_2)$ are isomorphic. $\square$

In this way, the set of all standard cyclic orders is countably categorical. Consequently as far as countable structures are concerned, there is only one standard cyclic order, the structure $(\mathbf{Q} \cup \{\infty\}, \prec)$ obtained from the structure $(\mathbf{Q}, <)$ of the rational numbers by the construction of section 2.3.

2.3 From Linear Orders to Cyclic Orders

Let $\mathcal{M} = (\mathcal{T}, <)$ be a linear order and ∞ be a point such that $\infty \notin \mathcal{T}$. Let $\mathcal{M}' = (\mathcal{T}', \prec')$ be the structure where $\mathcal{T}' = \mathcal{T} \cup \{\infty\}$ and $\prec'$ is the ternary relation on $\mathcal{T}'$ defined as follows for all $x, y, z \in \mathcal{T}'$:

- $\prec' (x, y, z)$ iff either $(x, y, z \in \mathcal{T} \text{ and } x < y < z)$ or $(x, y, z \in \mathcal{T} \text{ and } y < z < x)$ or $(x, y, z \in \mathcal{T} \text{ and } z < x < y)$ or $(x = \infty, y, z \in \mathcal{T} \text{ and } y < z)$ or $(y = \infty, x, z \in \mathcal{T} \text{ and } z < x)$ or $(z = \infty, x, y \in \mathcal{T} \text{ and } x < y)$.

$\mathcal{M}' = (\mathcal{T}', \prec')$ is called cyclic order on $\mathcal{M} = (\mathcal{T}, <)$ and ∞ . The reader may easily verify the following result.

Proposition 3. *Let $\mathcal{M} = (\mathcal{T}, <)$ be a linear order and ∞ be a point such that $\infty \notin \mathcal{T}$. The cyclic order $\mathcal{M}' = (\mathcal{T}', \prec')$ on $\mathcal{M} = (\mathcal{T}, <)$ and ∞ is a cyclic order. Moreover, if $\mathcal{M} = (\mathcal{T}, <)$ is dense and unbounded then $\mathcal{M}' = (\mathcal{T}', \prec')$ is standard.*

2.4 From Cyclic Orders to Linear Orders

Let $\mathcal{M} = (\mathcal{T}, \prec)$ be a cyclic order and ∞ be a point such that $\infty \in \mathcal{T}$. Let $\mathcal{M}' = (\mathcal{T}', <')$ be the structure where $\mathcal{T}' = \mathcal{T} \setminus \{\infty\}$ and $<'$ is the binary relation on $\mathcal{T}'$ defined as follows for all $x, y \in \mathcal{T}'$:

- $x <' y$ iff $\prec (x, y, \infty)$.

$\mathcal{M}' = (\mathcal{T}', <')$ is called linear order on $\mathcal{M} = (\mathcal{T}, \prec)$ and ∞ . The reader may easily verify the following result.

Proposition 4. *Let $\mathcal{M} = (\mathcal{T}, \prec)$ be a cyclic order and ∞ be a point such that $\infty \in \mathcal{T}$. The linear order $\mathcal{M}' = (\mathcal{T}', <')$ on $\mathcal{M} = (\mathcal{T}, \prec)$ and ∞ is a linear order. Moreover, if $\mathcal{M} = (\mathcal{T}, \prec)$ is standard then $\mathcal{M}' = (\mathcal{T}', <')$ is dense and unbounded.*

2.5 An Equivalence Result

We now prove that the two constructions defined above yield an equivalence result between the two classes of models. Let $\mathcal{M} = (\mathcal{T}, <)$ be a countable linear order, ∞ be a point such that $\infty \notin \mathcal{T}$. Let $\mathcal{M}' = (\mathcal{T}', \prec')$ be the cyclic order on $\mathcal{M} = (\mathcal{T}, <)$ and ∞ ; let ∞' be a point such that $\infty' \in \mathcal{T}'$ and $\mathcal{M}'' = (\mathcal{T}'', <'')$ be the linear order on $\mathcal{M}' = (\mathcal{T}', \prec')$ and ∞' . Let us assume that $\mathcal{M} = (\mathcal{T}, <)$ is dense and unbounded. If $\infty = \infty'$ then it is easily seen that the function f with domain $\mathcal{T}$ and range $\mathcal{T}''$ defined as follows for all $x \in \mathcal{T}$:

- $f(x) = x$;

is an isomorphism from $\mathcal{M} = (\mathcal{T}, <)$ to $\mathcal{M}'' = (\mathcal{T}'', <'')$. If $\infty \neq \infty'$ then the sub-models $\mathcal{M}^- = (\mathcal{T}^-, <^-)$ and $\mathcal{M}^+ = (\mathcal{T}^+, <^+)$ where $\mathcal{T}^- = \{x: x < \infty'\}$ and $\mathcal{T}^+ = \{x: \infty' < x\}$ are dense and unbounded. By proposition 1, $\mathcal{M}^- = (\mathcal{T}^-, <^-)$ and $\mathcal{M}^+ = (\mathcal{T}^+, <^+)$ are isomorphic. Hence there is a function $f^\mp$ with domain $\mathcal{T}^-$ and range $\mathcal{T}^+$ such that $f^\mp$ is an isomorphism from $\mathcal{M}^- = (\mathcal{T}^-, <^-)$ to $\mathcal{M}^+ = (\mathcal{T}^+, <^+)$ and there is a function $f^\pm$ with domain $\mathcal{T}^+$ and range $\mathcal{T}^-$ such that $f^\pm$ is an isomorphism from $\mathcal{M}^+ = (\mathcal{T}^+, <^+)$ to $\mathcal{M}^- = (\mathcal{T}^-, <^-)$. Therefore it is easily seen that the function f with domain $\mathcal{T}$ and range $\mathcal{T}''$ defined as follows for all $x \in \mathcal{T}$:

- If $x = \infty'$ then $f(x) = \infty$;
- If $x < \infty'$ then $f(x) = f^\mp(x)$;

- If $\infty' < x$ then $f(x) = f^\pm(x)$;

is an isomorphism from $\mathcal{M} = (\mathcal{T}, <)$ to $\mathcal{M}'' = (\mathcal{T}'', <'')$. Hence the following result.

Proposition 5. *Let $\mathcal{M} = (\mathcal{T}, <)$ be a countable linear order, ∞ be a point such that $\infty \notin \mathcal{T}$, $\mathcal{M}' = (\mathcal{T}', <')$ be the cyclic order on $\mathcal{M} = (\mathcal{T}, <)$ and ∞, ∞' be a point such that $\infty' \notin \mathcal{T}'$ and $\mathcal{M}'' = (\mathcal{T}'', <'')$ be the linear order on $\mathcal{M}' = (\mathcal{T}', <')$ and ∞' . If $\mathcal{M} = (\mathcal{T}, <)$ is dense and unbounded then $\mathcal{M} = (\mathcal{T}, <)$ and $\mathcal{M}'' = (\mathcal{T}'', <'')$ are isomorphic.*

Let $\mathcal{M} = (\mathcal{T}, <)$ be a cyclic order, ∞ be a point such that $\infty \in \mathcal{T}$, $\mathcal{M}' = (\mathcal{T}', <')$ be the linear order on $\mathcal{M} = (\mathcal{T}, <)$ and ∞, ∞' be a point such that $\infty' \notin \mathcal{T}'$ and $\mathcal{M}'' = (\mathcal{T}'', <'')$ be the cyclic order on $\mathcal{M}' = (\mathcal{T}', <')$ and ∞' . It is easily seen that the function f with domain $\mathcal{T}$ and range $\mathcal{T}''$ defined as follows for all $x \in \mathcal{T}$:

- If $x \neq \infty$ then $f(x) = x$;
- If $x = \infty$ then $f(x) = \infty'$;

is an isomorphism from $\mathcal{M} = (\mathcal{T}, <)$ to $\mathcal{M}'' = (\mathcal{T}'', <'')$. Hence the following result.

Proposition 6. *Let $\mathcal{M} = (\mathcal{T}, <)$ be a cyclic order, ∞ be a point such that $\infty \in \mathcal{T}$, $\mathcal{M}' = (\mathcal{T}', <')$ be the linear order on $\mathcal{M} = (\mathcal{T}, <)$ and ∞, ∞' be a point such that $\infty' \notin \mathcal{T}'$ and $\mathcal{M}'' = (\mathcal{T}'', <'')$ be the cyclic order on $\mathcal{M}' = (\mathcal{T}', <')$ and ∞' . $\mathcal{M} = (\mathcal{T}, <)$ and $\mathcal{M}'' = (\mathcal{T}'', <'')$ are isomorphic.*

2.6 Elimination of Quantifiers

Let $\mathcal{L}_l$ be the first-order language consisting of the binary predicate $<$ and the binary predicate $=$. The theory Δ_l of dense and unbounded linear orders has 6 axioms:

- $(\forall x)(x \neq x)$
- $(\forall xyz)(x < y \wedge y < z \rightarrow x < z)$;
- $(\forall xy)(x = y \vee x < y \vee y < x)$;
- $(\forall xy)(x < y \rightarrow (\exists z)(x < z \wedge z < y))$;
- $(\forall x)(\exists y)(y < x)$;
- $(\forall x)(\exists y)(x < y)$.

By the Löwenheim-Skolem theorem, every Δ_l -consistent sentence ϕ in $\mathcal{L}_l$ has a countable model. By proposition 1, this model is isomorphic to the structure $(\mathbf{Q}, <)$ of the rational numbers. Hence for every sentence ϕ in $\mathcal{L}_l$, either ϕ is a consequence of Δ_l or $\neg\phi$ is a consequence of Δ_l . Consequently the set of consequences of Δ_l is maximal consistent and Δ_l is a complete theory. The method of elimination of quantifiers applies to Δ_l and gives a way of deciding whether or not a sentence ϕ in $\mathcal{L}_l$ is a consequence of Δ_l , see Langford [14]. It consists of proving the following result.

Proposition 7. *For every formula ϕ in $\mathcal{L}_l$ with free variables in $\{x, y_1, \dots, y_I\}$, the formula $(\exists x)\phi$ is Δ_l -equivalent to a Boolean combination of atomic formulas in $\mathcal{L}_l$ with free variables in $\{y_1, \dots, y_I\}$.*

Our aim is to define a similar method for the theory Δ_c of standard cyclic orders. Let $\mathcal{L}_c$ be the first-order language consisting of the ternary predicate $\prec$ and the binary predicate $=$. The theory Δ_c of standard cyclic orders has 6 axioms:

- $(\forall xy)(\neg \prec(x, y, y))$;
- $(\forall xyz)(\prec(x, y, z) \wedge \prec(x, z, t) \rightarrow \prec(x, y, t))$;
- $(\forall xyz)(x \neq y \wedge x \neq z \rightarrow y = z \vee \prec(x, y, z) \vee \prec(x, z, y))$;
- $(\forall xyz)(\prec(x, y, z) \leftrightarrow (y, z, x) \leftrightarrow (z, x, y))$;
- $(\forall xy)(x \neq y \rightarrow (\exists z) \prec(x, z, y))$;
- $(\forall xy)(x \neq y \rightarrow (\exists z) \prec(x, z, y))$.

Following the line of reasoning suggested above within the framework of dense linear orders, the reader may easily verify the following results:

- Every Δ_c -consistent sentence ϕ in $\mathcal{L}_c$ has a countable model;
- This model is isomorphic to the structure $(\mathbf{Q} \cup \{\infty\}, \prec)$ obtained from the structure $(\mathbf{Q}, <)$ of the rational numbers by the construction of section 2.3;
- For every sentence ϕ in $\mathcal{L}_c$, either ϕ is a consequence of Δ_c or $\neg\phi$ is a consequence of Δ_c ;
- The set of consequences of Δ_c is maximal consistent;
- Δ_c is a complete theory.

We now come to the method of elimination of quantifiers applied to the theory Δ_c . For our purpose, it suffices to prove that for every conjunction ϕ of the form $\prec(x, y_1, z_1) \wedge \dots \wedge \prec(x, y_I, z_I) \wedge \neg \prec(x, t_1, u_1) \wedge \dots \wedge \neg \prec(x, t_J, u_J) \wedge x = v_1 \wedge \dots \wedge x = v_K \wedge x \neq w_1 \wedge \dots \wedge x \neq w_L$, the formula $(\exists x)\phi$ is Δ_c -equivalent to a Boolean combination of atomic formulas with free variables in $\{y_1, \dots, y_I, z_1, \dots, z_I, t_1, \dots, t_J, u_1, \dots, u_J, v_1, \dots, v_K, w_1, \dots, w_L\}$. Firstly, it is easy to show that for every $i, i' \in \{1, \dots, I\}$, the formula $\prec(x, y_i, z_i) \wedge \prec(x, y_{i'}, z_{i'})$ is Δ_c -equivalent to a disjunction of formulas of the form $\prec(x, y, z) \wedge \phi''$ where $y, z \in \{y_i, y_{i'}, z_i, z_{i'}\}$ and ϕ'' is a Boolean combination of atomic formulas in $\mathcal{L}_c$ with free variables in $\{y_i, y_{i'}, z_i, z_{i'}\}$. Hence we may consider that $I = 0$ or $I = 1$. Secondly, we observe that for every $j \in \{1, \dots, J\}$, the formulas $\neg \prec(x, t_j, u_j)$ and $x = t_j \vee x = u_j \vee t_j = u_j \vee \prec(x, u_j, t_j)$ are Δ_c -equivalent. Consequently we may consider that $J = 0$. Thirdly, it should be clear that if $K \geq 1$ then the formula $(\exists x)\phi$ is Δ_c -equivalent to $\prec(v_1, y_1, z_1) \wedge \dots \wedge \prec(v_1, y_I, z_I) \wedge \neg \prec(v_1, t_1, u_1) \wedge \dots \wedge \neg \prec(v_1, t_J, u_J) \wedge v_1 = v_2 \wedge \dots \wedge v_1 = v_K \wedge v_1 \neq w_1 \wedge \dots \wedge v_1 \neq w_L$. Therefore let us assume that $K = 0$. Fourthly, the reader may check that for every $l, l' \in \{1, \dots, L\}$, the formulas $x \neq w_l \wedge x \neq w_{l'}$ and $(w_l = w_{l'} \wedge x \neq w_l) \vee (w_l = w_{l'} \wedge x \neq w_{l'}) \vee \prec(x, w_l, w_{l'}) \vee \prec(x, w_{l'}, w_l)$ are Δ_c -equivalent. Thus we may consider that $L = 0$ or $L = 1$. Since:

- the formulas $(\exists x)(\prec(x, y_1, z_1) \wedge x \neq w_1)$ and $y_1 \neq z_1$ are Δ_c -equivalent;
- the formulas $(\exists x)(\prec(x, y_1, z_1))$ and $y_1 \neq z_1$ are Δ_c -equivalent;
- the formulas $(\exists x)(x \neq w_1)$ and $\top$ are Δ_c -equivalent;

our proof of the following result is complete.

Proposition 8. *For every formula ϕ in $\mathcal{L}_c$ with free variables in $\{x, y_1, \dots, y_I\}$, the formula $(\exists x)\phi$ is Δ_c -equivalent to a Boolean combination of atomic formulas in $\mathcal{L}_c$ with free variables in $\{y_1, \dots, y_I\}$.*

3 The Cyclic Point Algebra

3.1 Entities and Relations

Let $\mathcal{C}$ be an oriented circle. The entities we consider are the points of this circle; we will denote them by v, w, x , *etc.*, and we will call them the cyclic points. Sometimes we will use a rational number belonging to the interval $[0, 360[$ to define a point of $\mathcal{C}$. Such a rational number expresses the angle from the horizontal line to the line passing through the center of $\mathcal{C}$ and intersecting the point of $\mathcal{C}$. Given two points $x, y \in \mathcal{C}$, $[x, y]$ denotes the set of the points found onto $\mathcal{C}$ by going from x to y by following the orientation of $\mathcal{C}$. The atomic relations considered between points of $\mathcal{C}$ are the six ternary relations defined in the following way :

$$\begin{aligned} B_{abc} &= \{(x, y, z) \in \mathcal{C}^3 : x \neq y, x \neq z, y \neq z \text{ and } y \in [x, z]\}, \\ B_{acb} &= \{(x, y, z) \in \mathcal{C}^3 : x \neq y, x \neq z, y \neq z \text{ and } z \in [x, y]\}, \\ B_{aab} &= \{(x, x, y) \in \mathcal{C}^3 : x \neq y\}, \\ B_{baa} &= \{(y, x, x) \in \mathcal{C}^3 : x \neq y\}, \\ B_{aba} &= \{(x, y, x) \in \mathcal{C}^3 : x \neq y\}, \\ B_{aaa} &= \{(x, x, x) \in \mathcal{C}^3\}. \end{aligned}$$

These six relations are illustrated in Fig. 3.1. We denote the set of these atomic relations by $\mathcal{B}_{\mathcal{C}}$ and in the sequel we will use a, b, c , *etc.* to designate them. We note that these atomic relations are complete and mutually exclusive, *i.e.* three cyclic points satisfy one, and only one, atomic relation of this set of qualitative relations. B_{abc} and B_{acb} correspond to both atomic relations satisfied in the cases where the three points are distinct points. B_{aab} , B_{baa} and B_{aba} are concerned with the cases where two of the three points are the same. The atomic relation B_{aaa} corresponds to the case in which the three points are equal. From the atomic relations of $\mathcal{B}_{\mathcal{C}}$ we define the set of the complex relations of

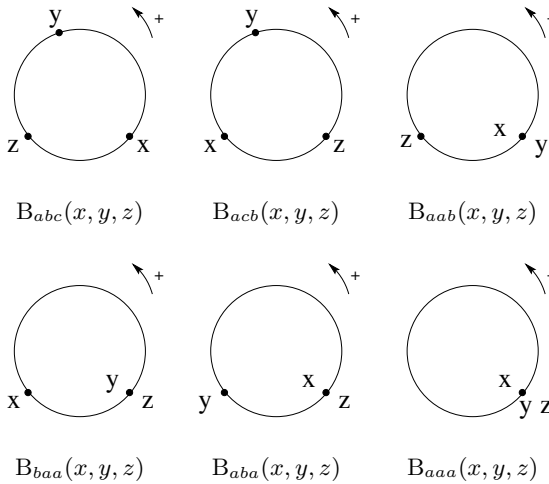


Fig. 1. Atomic relations of $\mathcal{B}_{\mathcal{C}}$.

the cyclic point algebra by taking the subsets of $\mathcal{B}_C$, *i.e.* $2^{\mathcal{B}_C}$. In the sequel we will say relation for complex relation. α, β, γ , *etc.*, will denote the relations. We have a set of $2^6 = 64$ relations with two particular relations : the empty relation $\{\}$ (also denoted by $\emptyset$) and the total relation $\{B_{aaa}, B_{aab}, B_{baa}, B_{aba}, B_{abc}, B_{acb}\}$ (also improperly denoted by $2^{\mathcal{B}_C}$). Given a relation $\alpha \in 2^{\mathcal{B}_C}$ and three cyclic points x, y, z , we have $\alpha(x, y, z)$ if, and only if, there exists $a \in \alpha$ such that $a(x, y, z)$. Such a relation α can be seen as the disjunction of its atomic relations. With relations of $\mathcal{B}_C$ we can represent incomplete information about relative positions of cyclic points.

3.2 Basic Operations

The binary operations of intersection $\cap$, union $\cup$ and the unary operation of complement are defined onto $2^{\mathcal{B}_C}$ in the following way :

$$\begin{aligned} \forall x, y, z \in \mathcal{C}, \forall \alpha, \beta \in 2^{\mathcal{B}_C}, \\ (\alpha \cup \beta)(x, y, z) &\leftrightarrow \alpha(x, y, z) \text{ or } \beta(x, y, z), \\ (\alpha \cap \beta)(x, y, z) &\leftrightarrow \alpha(x, y, z) \text{ and } \beta(x, y, z), \\ \bar{\alpha}(x, y, z) &\leftrightarrow \text{not } \alpha(x, y, z). \end{aligned}$$

These operations can be seen as the usual set operations since we have :

$$\begin{aligned} \forall a \in \mathcal{B}_C, \forall \alpha, \beta \in 2^{\mathcal{B}_C}, \\ a \in (\alpha \cup \beta) &\leftrightarrow a \in \alpha \text{ or } a \in \beta, \\ a \in (\alpha \cap \beta) &\leftrightarrow a \in \alpha \text{ and } a \in \beta, \\ a \in \bar{\alpha} &\leftrightarrow a \notin \alpha. \end{aligned}$$

We also define the unary operations of permutation (denoted by $\curvearrowright$) and of rotation (denoted by $\curvearrowleft$) in the following way:

$$\begin{aligned} \text{Let } a \in \mathcal{B}_C, \\ \forall x, y, z \quad a^{\curvearrowright}(x, z, y) &\leftrightarrow a(x, y, z), \\ \forall x, y, z \quad a^{\curvearrowleft}(y, z, x) &\leftrightarrow a(x, y, z). \end{aligned}$$

Table 1. Permutations and rotations of the atomic relations of $\mathcal{B}_C$.

a	B_{aaa}	B_{aab}	B_{baa}	B_{aba}	B_{abc}	B_{acb}
$a^{\curvearrowright}$	B_{aaa}	B_{aba}	B_{baa}	B_{aab}	B_{acb}	B_{abc}
$a^{\curvearrowleft}$	B_{aaa}	B_{aba}	B_{aab}	B_{baa}	B_{abc}	B_{acb}

Table 1 gives the permutation and the rotation of each atomic relation. We extend these operations to the relations of $2^{\mathcal{B}_C}$. The permutation (resp. the rotation) of $\alpha \in 2^{\mathcal{B}_C}$, denoted by $\alpha^{\curvearrowright}$ (resp. by $\alpha^{\curvearrowleft}$), is the union of the permutations (resp. the rotations) of its atomic relations. Given four cyclic points w, x, y, z , from the atomic relation a satisfied by w, x, y and the atomic relation b satisfied by x, y, z , we can deduce the possible atomic relations of $\mathcal{B}_C$ satisfied by w, x, z . This set of atomic relations is given by the binary

operation of composition, which is denoted by $\circ$. More formally, $a \circ b$ is the relation of $2^{\mathcal{B}_C}$ defined by:

$$a \circ b = \{c \in \mathcal{B}_C : \exists w, x, y, z \ a(w, x, y) \ \& \ b(x, y, z) \ \& \ c(w, x, z)\}.$$

The composition of two relations α and β of $2^{\mathcal{B}_C}$ is defined by $\alpha \circ \beta = \bigcup_{a \in \alpha, b \in \beta} a \circ b$. In the sequel we will suppose that the unary operations have priority. Knowing the relations α and β satisfied respectively by three cyclic points w, x, y and the three points x, y, z , with the operation of composition we can deduce the possible atomic relations satisfied by w, x, z . Let us note that with the operations of composition, permutation and rotation we can also find the set of atomic relations which can be satisfied by w, y, z . This set of atomic relations corresponds to the relation $\alpha^{\curvearrowright} \circ \beta^{\curvearrowright}$.

Table 2. Table of composition for the atomic relations of $\mathcal{B}_C$.

$\circ \uparrow$	B_{aaa}	B_{aab}	B_{baa}	B_{aba}	B_{abc}	B_{acb}
B_{aaa}	$\{B_{aaa}\}$	$\{B_{aab}\}$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
B_{aab}	$\emptyset$	$\emptyset$	$\{B_{aab}\}$	$\{B_{aaa}\}$	$\{B_{aab}\}$	$\{B_{aab}\}$
B_{baa}	$\{B_{baa}\}$	$\{B_{aba}, B_{abc}, B_{acb}\}$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
B_{aba}	$\emptyset$	$\emptyset$	$\{B_{aba}\}$	$\{B_{baa}\}$	$\{B_{acb}\}$	$\{B_{abc}\}$
B_{abc}	$\emptyset$	$\emptyset$	$\{B_{abc}\}$	$\{B_{baa}\}$	$\{B_{aba}, B_{abc}, B_{acb}\}$	$\{B_{abc}\}$
B_{acb}	$\emptyset$	$\emptyset$	$\{B_{acb}\}$	$\{B_{baa}\}$	$\{B_{acb}\}$	$\{B_{aba}, B_{abc}, B_{acb}\}$

3.3 Constraint Networks of Cyclic Points

To represent spatial information about cyclic points we use particular ternary constraint networks which we call constraint networks of cyclic points, CNCP in short. Each variable of a CNCP represents a cyclic point and each ternary constraint is defined by a relation of $2^{\mathcal{B}_C}$. This relation represents all allowed relative positions satisfied by the three points represented by the three concerned variables. More formally, a CNCP is defined in the following way.

Definition 9. A CNCP is a pair $\mathcal{N} = (V, C)$, where:

- V is a nonempty set of n ordered variables $\{V_1, \dots, V_n\}$ representing n cyclic points.
- D is a mapping from $(V \times V \times V)$ to $2^{\mathcal{B}_C}$ defining the constraints onto V . In the sequel, usually we will denote $C(V_i, V_j, V_k)$ by C_{ijk} or $C_{v_i v_j v_k}$.

With no loss of generality, we may suppose that each CNCP satisfies the following properties:

- for all $i, j, k \in \{1, \dots, n\}$, $C_{ijk} = C_{kij} = C_{ikj}^{\curvearrowright}$.
- For all $i, j \in \{1, \dots, n\}$, $C_{iiij} \subseteq \{B_{aab}, B_{aaa}\}$,
- For all $i \in \{1, \dots, n\}$, $C_{iii} \subseteq \{B_{aaa}\}$.

Intuitively, the requirement (a) stipulates that constraints between three variables must be coherent. Consequently, giving the constraints C_{ijk} for all $i, j, k \in \{1, \dots, n\}$ with $i \leq j \leq k$ is sufficient to define a CNCP. Concerning conditions (b) and (c), let us note that for convenience reasons we allow the empty relation for C_{iij} and C_{iii} . The second and third conditions contain necessary atomic relations allowing one or many equalities of cyclic points.

Given a CNCP, an important issue is to determine its consistency, *i.e.* whether there exists a set of cyclic points satisfying its constraints. More formally, given a CNCP $\mathcal{N} = (V, C)$, we have the following definitions:

- An instantiation m of $\mathcal{N}$ is a function from V to $\mathcal{C}$ associating with each variable $V_i \in V$ a cyclic point $m(V_i)$ (with $i \in \{1, \dots, |V|\}$). In what follows, $m(V_i)$ will be sometimes denoted by m_i , or m_{V_i} . m_{ijk} (equally denoted by $m_{V_i V_j V_k}$ or $m(V_i, V_j, V_k)$) is the atomic relation of $\mathcal{B}_{\mathcal{C}}$ satisfied by the points m_i, m_j and m_k with $i, j, k \in \{1, \dots, n\}$ ($n = |V|$).
- An instantiation m of $\mathcal{N}$ is consistent iff for all $i, j, k \in \{1, \dots, n\}$, $m_{ijk} \subseteq C_{ijk}$ ($n = |V|$). we will say that m is a solution of $\mathcal{N}$.
- A partial instantiation m of $\mathcal{N}$ is a mapping from V' to $\mathcal{C}$, with $V' \subseteq V$, which associates with each variable of V' a cyclic point. m is a partial solution of $\mathcal{N}$ iff the points associated with the variables of V' satisfy the ternary constraints uniquely concerning by the variables of V' .
- $\mathcal{N}$ is consistent iff it admits a consistent instantiation.

To solve the consistency problem of a qualitative binary constraint network, the path-consistency method is usually used. This method consists of obtaining a constraint network equivalent to the initial network (a network with exactly the same solutions) by deleting some atomic relations which do not participate in any solution. This method uses the operations of composition, inverse and intersection. The obtained network is 3-consistent, *i.e.* we can always extend a partial solution concerning two variables to a partial solution concerning a third variable in addition to the first ones. In a similar way, we can use the operations of composition, intersection, rotation and permutation to remove impossible atomic relations in the constraints of a CNPC. In particular we can apply the following operations onto a CNPC $\mathcal{N} = (V, C)$:

$$\begin{aligned} C_{ijk} &\leftarrow C_{ijk} \cap (C_{ijl} \circ C_{jlk}), \\ C_{jki} &\leftarrow C_{ijk}^{\sim}, C_{kij} \leftarrow C_{jki}^{\sim}, \\ C_{ikj} &\leftarrow C_{ijk}^{\leftrightarrow}, C_{jik} \leftarrow C_{jki}^{\leftrightarrow}, C_{kji} \leftarrow C_{kij}^{\leftrightarrow}. \end{aligned}$$

onto all each 4-tuple $i, j, k, l \in \{1, \dots, |V|\}$ until a fixed point is reached. This method is accomplished in polynomial time, we will call it the composition closure method. The obtained CNCP admits the same solutions as the initial CNCP. Moreover this constraint network is closed for composition: we say that a CNCP $\mathcal{N} = (V, C)$ is closed for the operation of composition iff

$$\forall i, j, k, l \in \{1, \dots, |V|\}, \quad C_{ijk} \subseteq C_{ijl} \circ C_{jlk}.$$

To close this subsection let us note that for all $i, j, k, l \in \{1, \dots, |V|\}$, if $C_{ijk} \subseteq C_{ijl} \circ C_{jlk}$ then the inclusion $C_{jki} \subseteq C_{jkl} \circ C_{kli}$ and the inclusion $C_{ikj} \subseteq C_{ikl} \circ C_{klj}$ are not always satisfied.

3.4 Tractability Results about CNCP

In the sequel we will prove that the consistency problem for the CNCP, which we will denote C-CNCP, is NP-complete in the general case. This subsection is devoted to the definitions of tractable cases and intractable cases. These cases will be parametrized by two nonempty sets B and T , where B is the union of $\{\emptyset\}$ and zero or several sets among the following ones : $\{\{B_{aaa}\}\}$, $\{\{B_{aab}\}\}$, $\{B_{aba}\}$, $\{B_{baa}\}$, $\{\{B_{aaa}, B_{aab}\}\}$, $\{B_{aaa}, B_{aba}\}$, $\{B_{aaa}, B_{baa}\}$, and where T is a subset of 2^{B^c} containing the empty relation. B contains the possible constraints between two distinct variables, whereas T contains the possible constraints between three distinct variables. In other words, B corresponds to the relations C_{iij} and T to the relations C_{ijk} , with i, j, k three distinct integers. T must be closed for the operations of permutation and composition. $\text{CNCP}(B, T)$ will denote the set of the CNCP $\mathcal{N} = (V, C)$ such that for all $i, j \in \{1, \dots, n\}$ (with $n = |V|$) if $i \neq j$ then $C_{iij} \in B$ and for all $i, j, k \in \{1, \dots, n\}$ if i, j, k are pairwise distinct then $C_{ijk} \in T$. $\text{C-CNCP}(B, T)$ will denote the consistency problem restricted to the networks of $\text{CNCP}(B, T)$.

Table 3. The sets B_0, B_1, B_2 and B_3 .

	$\emptyset$	$\{B_{aaa}\}$	$\{B_{aab}\}, \{B_{aba}\}, \{B_{baa}\}$	$\{B_{aaa}, B_{aab}\}, \{B_{aaa}, B_{aba}\}, \{B_{aaa}, B_{baa}\}$
B_0	• •			
B_1	•		•	
B_2	•		•	•
B_3	• •		•	•
B_4	• •	•		

Let us note that we have eight possible different sets B . The sets considered for B in the sequel are given in Table 3 and those for T are defined in Table 5. Firstly we define tractable cases and secondly, we give some intractable cases. Let us start our study with an easy case.

Proposition 10. $\text{C-CNCP}(B_0, T)$ is a polynomial problem.

Proof. Let $\mathcal{N} = (V, C) \in \text{CNCP}(B_0, T)$. $\mathcal{N}$ is a consistent network iff for all $i, j, k \in \{1, \dots, |V|\}$, $B_{aaa} \in C_{ijk}$. This test can be established in time $O(n^3)$ with $n = |V|$. $\square$

Proposition 11. Let B be a set of relations such that $B_1 \subseteq B$ and let T be a set closed for the operation of intersection (in addition to the closure for the operations of permutation and rotation). Let $T' = (T \setminus B_3) \cup \{\emptyset\}$ and $B' = (B \setminus B_0) \cup \{\emptyset\}$. $\text{C-CNCP}(B, T)$ is a polynomial problem (resp. a NP-complete problem) iff $\text{C-CNCP}(B', T')$ is a polynomial problem (resp. a NP-complete problem).

Proof. Trivially, since $B' \subseteq B$ and $T' \subseteq T$ if $\text{C-CNCP}(B, T)$ is a polynomial problem then $\text{C-CNCP}(B', T')$ is also a polynomial problem and if $\text{C-CNCP}(B', T')$ is a NP-complete complete then $\text{C-CNCP}(B, T)$ is equally a NP-complete problem. Now, let us define a polynomial transformation from $\text{C-CNCP}(B, T)$ to $\text{C-CNCP}(B', T')$. Let

$\mathcal{N} = (V, C)$ be a CNCP belonging to $\text{CNCP}(B, T)$. From $\mathcal{N}$ we define a CNCP $\mathcal{N}''$ by making the following steps.

Step 1. We initialize $\mathcal{N}' = (V', C')$ by $\mathcal{N}$.

Step 2. We define the binary graph $G = (S, E)$ in the following way : for each variable $V_i' \in V'$ there is an associated node $s_i \in S$ and (s_i, s_j) belongs to the set of edges E iff $C'_{ijk} \subseteq \{B_{aaa}, B_{aab}\}$ for some $k \in \{1, \dots, |V'|\}$. Let $C_1, \dots, C_p$ be the strongly connected components of G .

Step 3. We define the CNCP $\mathcal{N}'' = (V'', C'')$ by:

- with each component C_i is associated a variable V_i'' for each $i \in \{1, \dots, p\}$.
- $C''_{ijk} = \bigcap_{s_r \in C_i, s_s \in C_j, s_t \in C_k} C'_{rst}$ for each $i, j, k \in \{1, \dots, p\}$.

Step 4. If $\mathcal{N}'$ and $\mathcal{N}''$ are identical then we stop. If not we set $\mathcal{N}'$ to $\mathcal{N}''$ and we go back to Step 2.

From the construction of $\mathcal{N}''$ we note that $\mathcal{N}''$ is consistent iff $\mathcal{N}$ is consistent. Since T is closed for the operation of intersection we know that C''_{ijk} belongs to T for all distinct integers i, j, k . Since the relations of B contain at most two atomic relations and $B_1 \subseteq B$ it follows that $C''_{ijj} \in B \cup \{B_{aaa}\}$. Moreover, by construction C''_{ijk} cannot belong to the set $B_3 \setminus \{\emptyset\}$ and C''_{iij} cannot be $\{B_{aaa}\}$. It follows that $\mathcal{N}'' \in \text{CNCP}(B', T')$. The construction of $\mathcal{N}''$ is realized in polynomial time. It follows that if $\text{C-CNCP}(B', T')$ is a polynomial problem then $\text{C-CNCP}(B, T)$ is also a polynomial problem and if $\text{C-CNCP}(B, T)$ is a NP-complete problem then $\text{C-CNCP}(B', T')$ is also a NP-complete problem. $\square$

Lemma 12. *The composition closure method solves $\text{C-CNCP}(B_2, \{\{B_{abc}\}, \{B_{acb}\}, \emptyset\})$.*

Proof. Let $\mathcal{N}' = (V, C') \in \text{C-CNCP}(B_2, \{\{B_{abc}\}, \{B_{acb}\}, \emptyset\})$ and let $\mathcal{N} = (V, C)$ be the network obtained by applying the composition closure method on $\mathcal{N}'$. Let us show that $\mathcal{N} \in \text{CNCP}(B_1, \{\{B_{abc}\}, \{B_{acb}\}, \emptyset\})$. Let $i, j \in \{1, \dots, n\}$ with $n = |V|$ and $i \neq j$. We will suppose that $n > 3$ since the case $n \leq 3$ is trivial. We now prove that $C_{iij} \neq \{B_{aaa}\}$ and $C_{iij} \neq \{B_{aaa}, B_{aab}\}$. Let us suppose the contrary; since $\mathcal{N}$ is closed for composition it follows that $C_{iij} \subseteq C_{iik} \circ C_{ikj}$ for all $k \in \{1, \dots, n\}$. In the case where k is distinct from i and j we have $C_{ijk} = \{B_{abc}\}$ or $C_{ijk} = \{B_{acb}\}$. It follows that $C_{iij} \subseteq \{B_{aaa}, B_{aab}\} \circ \{B_{acb}, B_{abc}\}$. Consequently, $C_{iij} \subseteq \{B_{aab}\}$, which is a contradiction. We can conclude that $\mathcal{N} \in \text{CNCP}(B_1, \{\{B_{abc}\}, \{B_{acb}\}\})$.

Let us suppose that $\mathcal{N}$ does not contain the empty relation. We are going to construct a consistent instantiation m for $\mathcal{N}$ in which two distinct variables are associated with two distinct cyclic points. It is always possible to instantiate the first three variables. Now, let us suppose that we have a partial solution $m_1, \dots, m_{q-1}$ with $q > 3$ and $q \leq n$, such that $m_i \neq m_j$ for all $i, j \in \{1, \dots, q-1\}$. Let us show that we can extend this partial solution to the variable V_q by a cyclic point different from the other cyclic points used. Firstly, we renumber the variables $V_1, \dots, V_{q-1}$ such that $m_{i(i+1)(i+2)} = B_{abc}$ for each $i \in \{1, \dots, q-3\}$ and $m_{(q-2)(q-1)1} = B_{abc}$. Let $l \in \{1, \dots, q-1\}$ be such that $C_{lq(l \bmod (q-1)+1)} = \{B_{abc}\}$. Let us show the existence of l . Let us suppose that l does not exist. Hence we have for each $l \in \{1, \dots, q-2\}$, $C_{lq(l+1)} = \{B_{acb}\}$ and $C_{(q-1)q1} = \{B_{acb}\}$. Since $\mathcal{N}$ is closed for the operation of composition we have

$C_{1q3} \subseteq C_{1q2} \circ C_{q23}$. Consequently, $C_{1q3} \subseteq \{B_{acb}\} \circ \{B_{abc}\}$. Since this last composition equals $\{B_{acb}\}$ we can deduce that $C_{1q3} = \{B_{acb}\}$. By propagation and by using the more general fact that $C_{1qi} \subseteq C_{1q(i-1)} \circ C_{q(i-1)i}$ for each $i \in \{3, \dots, q-1\}$ we obtain the following equality: $C_{1q(q-1)} = \{B_{acb}\}$, hence by permutation we obtain $C_{1(q-1)q} = \{B_{abc}\}$. By rotation we have $C_{(q-1)q1} = \{B_{abc}\}$, which is a contradiction. We can conclude that there exists an integer l satisfying the given conditions. By defining m_q by a cyclic point such that $m_{lq(l \bmod (q-1)+1)}$ is B_{abc} (i.e. any intermediate cyclic point between m_l and $m_{l \bmod (q-1)+1}$ by following the circle orientation we extend m to a partial instantiation such that the valuations are pairwise distinct. Let us show that m is always a partial solution. In order to do that, let us suppose that there exists $i, j \in \{1, \dots, q-1\}$ such that $m_{ijq} \neq C_{ijq}$. i, j, q must be pairwise distinct. With no loss of generality we may suppose that $i < j$; then three cases are possible:

- $i < j \leq l$. We have $m_{ijq} = B_{abc}$, consequently $C_{ijq} = \{B_{acb}\}$. If $j = l$ then $C_{ilq} = \{B_{acb}\}$. In the contrary case, since $C_{liq} \subseteq C_{lij} \circ C_{ijq}$, it follows that $C_{liq} \subseteq \{B_{abc}\} \circ \{B_{acb}\}$ and hence $C_{liq} = \{B_{abc}\}$. Consequently $C_{ilq} = \{B_{acb}\}$. If $i = 1$ and $l = q-1$ then $C_{lqi} = \{B_{abc}\}$ and hence $C_{ilq} = \{B_{abc}\}$, which is a contradiction. In the contrary case $C_{ilq} \subseteq C_{il(l \bmod (q-1)+1)} \circ C_{l(l \bmod (q-1)+1)q}$ with $C_{il(l \bmod (q-1)+1)} \circ C_{l(l \bmod (q-1)+1)q} = \{B_{abc}\} \circ \{B_{acb}\} = \{B_{abc}\}$. Consequently $C_{ilq} = \{B_{abc}\}$, which is a contradiction.
- $l \leq i < j$.
 - Let us consider the case where $j = l+1$. It follows that $i = l$. Hence $m_{ijq} = B_{acb}$ and by consequence $C_{ijq} = \{B_{abc}\}$. This is a contradiction since $C_{l(l+1)q} = \{B_{abc}\}$.
 - Let us consider the case where $i = l+1$. It follows that $m_{ijq} = B_{abc}$ and then $C_{ijq} = \{B_{acb}\}$. It follows that $C_{(l+1)jq} = \{B_{acb}\}$. We know that $C_{j(l+1)l} \subseteq C_{j(l+1)q} \circ C_{(l+1)ql}$. Hence, $C_{j(l+1)l} \subseteq \{B_{abc}\} \circ \{B_{acb}\}$ and $C_{j(l+1)l} = \{B_{abc}\}$. This is a contradiction since $m_{j(l+1)l} = B_{acb}$.
 - Let us consider the case where $i \neq l+1$ and $j \neq l+1$. By propagating the fact that $C_{q(l+1)(q-m-2)} \subseteq C_{q(l+1)(q-m-1)} \circ C_{(l+1)(q-m-1)(q-m-2)}$ for $m \in \{0, \dots, q-2-j\}$. We obtain $C_{q(l+1)(q-m-2)} \subseteq \{B_{abc}\} \circ \{B_{acb}\}$ and hence $C_{q(l+1)(q-m-2)} \subseteq \{B_{abc}\}$, for $m \in \{1, \dots, q-2-j\}$. It follows that $C_{q(l+1)j} = \{B_{abc}\}$. Since $C_{qji} \subseteq C_{qj(l+1)} \circ C_{j(l+1)i}$. It results that $C_{qji} \subseteq \{B_{abc}\} \circ \{B_{acb}\}$ and thus $C_{qji} = \{B_{abc}\}$, which is a contradiction.
- $i < l < j$. Consequently we have $m_{ijq} = B_{acb}$ and then $C_{ijq} = \{B_{abc}\}$. Since $C_{i(l+1)q} \subseteq C_{i(l+1)l} \circ C_{(l+1)lq}$, we obtain $C_{i(l+1)q} \subseteq \{B_{abc}\} \circ \{B_{abc}\}$ and thus $C_{i(l+1)q} = \{B_{abc}\}$. If $j = l+1$ then we get a contradiction. Let us suppose that $j \neq l+1$. As $C_{qij} \subseteq C_{qi(l+1)} \circ C_{i(l+1)j}$ it follows that $C_{qij} \subseteq \{B_{abc}\} \circ \{B_{abc}\}$ and thus $C_{qij} = \{B_{abc}\}$. By rotation we have $C_{ijq} = \{B_{acb}\}$, which is a contradiction.

□

Proposition 13. Let T_0 be the set $B_3 \cup \{\{B_{abc}\}, \{B_{acb}\}, \emptyset\}$. $C\text{-CNCP}(B_3, T_0)$ is a polynomial problem.

Proof. From Lemma 12 it follows that $C\text{-CNCP}(B_2, \{\{B_{abc}\}, \{B_{acb}\}, \emptyset\})$ is a polynomial problem. From Proposition 11 we can conclude that $C\text{-CNCP}(B_2 \cup B_0, \{\{B_{abc}\},$

$\{B_{acb}\} \cup B_3$), i.e. $C\text{-CNCP}(B_3, T_0)$, is also a polynomial problem since T_0 is closed for the operation of intersection. $\square$

Proposition 14. *Let T_1 be a set composed of all relations minus the relations containing both atomic relations $\{B_{abc}\}$ and $\{B_{acb}\}$. $C\text{-CNCP}(B_4, T_1)$ is a polynomial problem.*

Proof. The set T_1 is closed for the operation of intersection and B_1 is a subset of B_4 , from Proposition 11 it follows that $C\text{-CNCP}(B_4, T_1)$ is a polynomial problem if, and only if, $C\text{-CNCP}(B_1, T_1 \setminus B_3)$ is a polynomial problem. Let $\mathcal{N} = (V, C) \in \text{CNCP}(B_1, T_1 \setminus B_3)$ and a solution m de $\mathcal{N}$. For all $i, j, k \in \{1, \dots, n\}$ pairwise distinct, we have $m_{ijk} \in \{B_{abc}, B_{acb}\}$ (because of the possible constraints allowed by the set B_1). It follows that $\mathcal{N} = (V, C')$ admits the same solutions that the $\text{CNCP}(V, C')$ defined by $C'_{ijk} = C_{ijk} \setminus \{B_{aaa}, B_{aba}, B_{baa}, B_{aab}\}$ if i, j, k are pairwise distinct integers, $C'_{ijk} = C_{ijk}$ else, for $i, j, k \in \{1, \dots, n\}$. $(V, C') \in \text{CNCP}(B_1, \{B_{abc}, B_{acb}\})$. As $C\text{-CNCP}(B_1, \{B_{abc}, B_{acb}\})$ is a polynomial problem (Lemma 12) we can conclude that $C\text{-CNCP}(B_1, T_1)$ is also a polynomial problem. $\square$

Lemma 15. *Let T be the set $\{\{B_{abc}\}, \{B_{acb}\}, \{B_{aaa}\}, \{B_{abc}, B_{aaa}\}, \{B_{acb}, B_{aaa}\}, \emptyset\}$. $C\text{-CNCP}(B_3, T)$ is a polynomial problem.*

Proof. Let be $\mathcal{N}' = (V, C') \in \text{RCPC}(B_3, T)$. By applying the method of composition closure on $\mathcal{N}'$ we obtain $\mathcal{N} = (V, C)$ belonging also to $C\text{-CNCP}(B_3, T)$. Let us suppose that $\mathcal{N}$ does not contain the empty constraint. Let us show that $\mathcal{N}$ is consistent. Let i, j, k, l be four pairwise distinct integers belonging to the set $\{1, \dots, n\}$ with $n = |V|$. Let us suppose that C_{ijk} contains the atomic relation B_{aaa} . Since that $C_{ijk} \subseteq C_{ijl} \circ C_{jlk}$ and $C_{jik} \subseteq C_{jil} \circ C_{ilk}$ we can deduce that $B_{aaa} \in C_{ijl}, B_{aaa} \in C_{jlk}, B_{aaa} \in C_{jil}$ and $B_{aaa} \in C_{ilk}$. Consequently, a constraint on three distinct variables contains the atomic relation B_{aaa} if, and only if, every constraint on three distinct variables contains the atomic relation B_{aaa} . Let us suppose that for two distinct integers $i, j \in \{1, \dots, n\}$, C_{ij} does not contain the atomic relation B_{aaa} , i.e. $C_{ij} = \{B_{aab}\}$. Let $k \in \{1, \dots, n\}$ different from i and j . Let k be an integer different from i and j . We have $C_{ijk} \subseteq C_{ij} \circ C_{jik}$. Hence $C_{ijk} \subseteq \{B_{aba}\} \circ \{B_{aaa}, B_{abc}, B_{acb}\}$ and hence $C_{ijk} \subseteq \{B_{abc}, B_{acb}\}$. It follows that C_{ijk} does not contain the atomic relation B_{aaa} . We can conclude that either all constraints of $\mathcal{N}$ contain the atomic relation B_{aaa} , in which case $\mathcal{N}$ is consistent, or that $\mathcal{N}$ belongs $\text{RCPC}(B_3, T_0)$ and consequently deciding consistency is polynomial (Proposition 13). $\square$

Proposition 16. *Let T_3 be the set composed of the relations of B_3 and the relations of the set $\{\{B_{abc}\}, \{B_{acb}\}, \{B_{abc}, B_{aaa}\}, \{B_{acb}, B_{aaa}\}\}$. $C\text{-CNCP}(B_3, T_3)$ is a polynomial problem.*

Proof. T_3 is closed for the operation of intersection and moreover $B_0 \subseteq B_3$. It follows that $C\text{-CNCP}(B_3, T_3)$ is a polynomial problem iff $C\text{-CNCP}(B_2, (T_1 \setminus B_3) \cup \{\emptyset\})$ is a polynomial problem. This is actually the case by Lemma 15. $\square$

Proposition 17. *Let T_2 be the set composed of the relations of B_3 and all the relations including the relation $\{B_{abc}, B_{acb}\}$. $C\text{-CNCP}(B_3, T_2)$ is a polynomial problem.*

Proof. T_2 is a set closed for the operation of intersection, moreover $B_0 \subseteq B_3$. It follows that $C\text{-CNCP}(B_3, T_2)$ is a polynomial problem iff $C\text{-CNCP}(B_2, (T_2 \setminus B_3) \cup \{\emptyset\})$ is a polynomial problem. Let us show that $C\text{-CNCP}(B_2, T_2 \setminus B_3)$ is a polynomial problem. Let $\mathcal{N} \in \text{CNCP}(B_2, T_2 \setminus B_3)$. By giving to the variables pairwise distinct values we obtain a solution for $\mathcal{N}$ since $\{B_{abc}, B_{acb}\} \subseteq C_{ijk}$ for all i, j, k pairwise distinct integers and no constraint belonging to B_2 implies the equality of two variables. $\square$

Proposition 18. *$C\text{-CNCP}(B_3, T_4)$ is a polynomial problem.*

Proof. Let $\mathcal{N}$ be a CNCP belonging to $\text{CNCP}(B_3, T_4)$. Each consistent instantiation m of $\mathcal{N}$ uses only one or two distinct cyclic points. Consequently, we can reduce the domain of the variables to two distinct cyclic points u and v . From this fact we can define a polynomial transformation from $\text{CNCP}(B_3, T_4)$ to the 2-SAT problem in the following way. Let $\mathcal{N} = (V, C)$ be a CNCP belonging to $\text{CNCP}(B_3, T_4)$. We note $|V|$ by n . Let $L = \{l_1, \dots, l_n\}$ be a set of literals. For each $i \in \{1, \dots, n\}$, the variable V_i is associated with the literal l_i . Intuitively, l_i is true will correspond to the assignation u to V_i and l_i is false will correspond to the assignation v to V_i . We define a set of clauses $\mathcal{C}$ over L by adding to $\mathcal{C}$ a set of clauses s_{ijk} for each $i, j, k \in \{1, \dots, n\}$ in accordance with the values of C_{ijk} . This translation is given by Table 4.

Table 4. Translation of the cyclic constraints with clauses

the constraint C_{ijk}	the clause c_{ijk}
$\{B_{aaa}\}$	$(l_i \vee l_j) \wedge (l_i \vee \bar{l}_k) \wedge (l_j \vee \bar{l}_i) \wedge (l_j \vee \bar{l}_k) \wedge (l_k \vee \bar{l}_i) \wedge (l_k \vee \bar{l}_j)$
$\{B_{aab}\}$	$(l_i \vee l_j) \wedge (l_i \vee l_k) \wedge (\bar{l}_i \vee l_j) \wedge (\bar{l}_i \vee l_k) \wedge (\bar{l}_i \vee \bar{l}_k) \wedge (\bar{l}_j \vee \bar{l}_k)$
$\{B_{aba}\}$	c_{ikj}
$\{B_{baa}\}$	c_{jki}
$\{B_{aaa}, B_{aab}\}$	$(\bar{l}_i \vee l_j) \wedge (l_i \vee \bar{l}_j)$
$\{B_{aaa}, B_{aba}\}$	c_{ikj}
$\{B_{aaa}, B_{baa}\}$	c_{kji}
$\{B_{aaa}, B_{aab}, B_{aba}, B_{baa}\}$	$\emptyset$

Given m a solution of $\mathcal{N} = (V, C)$ we define the following interpretation $\mathcal{I}$ onto $\mathcal{L}$: for each $l_i \in L$, $\mathcal{I}(l_i) = \text{true}$ iff $m(V_i) = u$. The reader can verify that $\mathcal{I}$ is a model of all sets of clauses s_{ijk} . Now, from a model $\mathcal{I}$ of all sets of clauses s_{ijk} , a solution of $\mathcal{N}$ can be obtained by taking $m(V_i) = u$ iff $l_i \in L$, $\mathcal{I}(l_i) = \text{true}$, else $m(V_i) = v$, for each $V_i \in V$. It can be verified that m is a solution of $\mathcal{N}$. $\square$

Let us note that the relations $\{B_{aaa}, B_{aab}, B_{aba}\}$, $\{B_{aaa}, B_{aba}, B_{baa}\}$ and $\{B_{aaa}, B_{aab}, B_{baa}\}$ can be expressed with clauses containing three literals and cannot be expressed with clauses containing only two literals. Despite this fact, we cannot assert that adding these relations implies NP-completeness.

Lemma 19. *Let T be a set containing both relations $\{B_{abc}\}$, $\{B_{acb}\}$ and any relation α including $\{B_{abc}, B_{acb}\}$. $C\text{-CNCP}(B_1, T)$ is a NP-complete problem.*

Proof. We give a polynomial transformation from the *cyclic ordering* problem – a NP-complete referenced in [10] – to $C\text{-CNCP}(B_1, T)$. The *cyclic ordering* problem is as follows: given a set A and a set Tr of 3-tuples (e, f, g) with e, f, g three distinct elements belonging to A , answer the following question: is there a one-to-one function $\mathcal{F}$ which associates an integer belonging to $\{1, \dots, |A|\}$ with each element of A such that, for each 3-tuple $(e, d, f) \in Tr$ we have $\mathcal{F}(e) < \mathcal{F}(f) < \mathcal{F}(g)$ or $\mathcal{F}(g) < \mathcal{F}(e) < \mathcal{F}(f)$ or $\mathcal{F}(f) < \mathcal{F}(g) < \mathcal{F}(e)$.

Let (A, Tr) be an instance of the *cyclic ordering* problem. Let $\mathcal{N} = (V, C)$ be a CNCP defined in the following way:

- V is a set of $|A|$ variables. To each of these variables corresponds an element belonging to A (to two distinct variables corresponds two distinct elements). We denote by V_e the variable associated with the element $e \in A$.
- The constraints of C are defined as follows: given $e, f, g \in A$ which are pairwise distinct, if (e, f, g) or (f, g, e) or $(g, e, f) \in Tr$ then $C(V_e, V_f, V_g) = \{B_{abc}\}$, else if (f, e, g) or (e, g, f) or $(g, f, e) \in Tr$ then $C(V_e, V_f, V_g) = \{B_{acb}\}$, else $C(V_e, V_f, V_g) = \alpha$. $C(V_e, V_e, V_f) = \{B_{aab}\}$ and $C(V_e, V_e, V_e) = \{B_{aaa}\}$

Let us prove that there exists a one-to-one mapping $\mathcal{F}$ solution of (A, T) if, and only if, there exists a solution m of $\mathcal{N}$.

- Let $\mathcal{F}$ be a solution of (A, T) . Let P denote a cyclic point and O the center of the circle. We define an instantiation m of $\mathcal{N}$ by for each $e \in A$: $m(V_e)$ is the cyclic point such that the angle $(OP, Om(V_e))$ equals $(\mathcal{F}(e) - 1) \times (360/|A|)$ degrees. Then m is a solution of $\mathcal{N}$. Indeed, we note that if a 3-tuple (e, f, g) belongs to Tr then $\mathcal{F}(e) < \mathcal{F}(f) < \mathcal{F}(g)$ or $\mathcal{F}(g) < \mathcal{F}(e) < \mathcal{F}(f)$ or $\mathcal{F}(f) < \mathcal{F}(g) < \mathcal{F}(e)$. Consequently, by moving along the circle starting from P we first meet $m(V_e)$, then $m(V_f)$ and finally $m(V_g)$ or $m(V_g)$, then $m(V_e)$ and finally $m(V_f)$ or $m(V_f)$, then $m(V_g)$ and finally $m(V_e)$. It follows that $B_{abc}(m(V_e), m(V_f), m(V_g))$.
- Let us consider a consistent instantiation m of $\mathcal{N}$. Let us define the one-to-one function $\mathcal{F}$ from A to $\{1, \dots, |A|\}$ in the following way. Because of the constraints in C the cyclic points associated with the variables are pairwise distinct. Let P be a cyclic point. Let $e \in A$; $\mathcal{F}(e)$ is the cardinal of the set $\{v \in V : m(v) \in [P, m(V_e)]\}$. $\mathcal{F}$ is a one-to-one function, moreover $\mathcal{F}$ is a solution for (A, T) . Indeed, let us suppose that $(e, f, g) \in T$; it follows that $C(V_e, V_f, V_g) = \{B_{abc}\}$. Consequently, by moving along the circle starting from P we first meet $m(V_e)$, then $m(V_f)$ and finally $m(V_g)$, or $m(V_f)$, then $m(V_g)$ and finally $m(V_e)$ or $m(V_g)$, then $m(V_e)$ and finally $m(V_f)$. It follows that $\mathcal{F}(e) < \mathcal{F}(f) < \mathcal{F}(g)$ or $\mathcal{F}(g) < \mathcal{F}(e) < \mathcal{F}(f)$ or $\mathcal{F}(f) < \mathcal{F}(g) < \mathcal{F}(e)$. Hence $\mathcal{F}$ is a solution.

□

From the previous proposition we can deduce that in the general case $C\text{-CNCP}(B_2, 2^{\mathcal{B}_c})$ is a NP-complete problem.

4 Complexity and Conceptual Spaces

For a whole family of temporal and spatial calculi based on linear orderings, the complexity properties are closely related to geometrical and topological properties of the

Table 5. A listing of T_i sets.

Relations \ Sets	T_0	T_1	T_2	T_3	T_4	Relations \ Sets	T_0	T_1	T_2	T_3	T_4
$\emptyset$	•	•	•	•	•	$\{B_{aaa}, B_{baa}, B_{abc}\}$		•			
$\{B_{aaa}\}$	•	•	•	•	•	$\{B_{aaa}, B_{baa}, B_{acb}\}$		•			
$\{B_{aab}\}$	•	•	•	•	•	$\{B_{aaa}, B_{abc}, B_{acb}\}$			•		
$\{B_{aba}\}$	•	•	•	•	•	$\{B_{aab}, B_{aba}, B_{baa}\}$		•			
$\{B_{baa}\}$	•	•	•	•	•	$\{B_{aab}, B_{aba}, B_{abc}\}$		•			
$\{B_{abc}\}$	•	•		•		$\{B_{aab}, B_{aba}, B_{acb}\}$		•			
$\{B_{acb}\}$	•	•		•		$\{B_{aab}, B_{baa}, B_{abc}\}$		•			
$\{B_{aaa}, B_{aab}\}$	•	•	•	•	•	$\{B_{aab}, B_{baa}, B_{acb}\}$		•			
$\{B_{aaa}, B_{aba}\}$	•	•	•	•	•	$\{B_{aab}, B_{abc}, B_{acb}\}$			•		
$\{B_{aaa}, B_{baa}\}$	•	•	•	•	•	$\{B_{aba}, B_{baa}, B_{abc}\}$		•			
$\{B_{aaa}, B_{abc}\}$		•		•		$\{B_{aba}, B_{baa}, B_{acb}\}$		•			
$\{B_{aaa}, B_{acb}\}$		•		•		$\{B_{aba}, B_{abc}, B_{acb}\}$			•		
$\{B_{aab}, B_{aba}\}$		•			•	$\{B_{baa}, B_{abc}, B_{acb}\}$			•		
$\{B_{aab}, B_{baa}\}$		•			•	$\{B_{aaa}, B_{aab}, B_{aba}, B_{baa}\}$		•			•
$\{B_{aab}, B_{abc}\}$		•			•	$\{B_{aaa}, B_{aab}, B_{aba}, B_{abc}\}$		•			
$\{B_{aab}, B_{acb}\}$		•				$\{B_{aaa}, B_{aab}, B_{aba}, B_{acb}\}$		•			
$\{B_{aba}, B_{baa}\}$		•			•	$\{B_{aaa}, B_{aba}, B_{baa}, B_{abc}\}$		•			
$\{B_{aba}, B_{abc}\}$		•				$\{B_{aaa}, B_{aba}, B_{baa}, B_{acb}\}$		•			
$\{B_{aba}, B_{acb}\}$		•				$\{B_{aaa}, B_{baa}, B_{abc}, B_{acb}\}$			•		
$\{B_{baa}, B_{abc}\}$		•				$\{B_{aab}, B_{aba}, B_{baa}, B_{abc}\}$		•			
$\{B_{baa}, B_{acb}\}$		•				$\{B_{aab}, B_{aba}, B_{baa}, B_{acb}\}$		•			
$\{B_{abc}, B_{acb}\}$			•			$\{B_{aab}, B_{baa}, B_{abc}, B_{acb}\}$			•		
$\{B_{aaa}, B_{aab}, B_{aba}\}$	•					$\{B_{aba}, B_{baa}, B_{abc}, B_{acb}\}$			•		
$\{B_{aaa}, B_{aab}, B_{baa}\}$	•					$\{B_{aaa}, B_{aab}, B_{aba}, B_{baa}, B_{abc}\}$		•			
$\{B_{aaa}, B_{aab}, B_{abc}\}$	•					$\{B_{aaa}, B_{aab}, B_{aba}, B_{baa}, B_{acb}\}$		•			
$\{B_{aaa}, B_{aab}, B_{acb}\}$	•					$\{B_{aaa}, B_{aba}, B_{baa}, B_{abc}, B_{acb}\}$			•		
$\{B_{aaa}, B_{aba}, B_{baa}\}$	•					$\{B_{aab}, B_{aba}, B_{baa}, B_{abc}, B_{acb}\}$			•		
$\{B_{aaa}, B_{aba}, B_{abc}\}$	•					$\{B_{aaa}, B_{aab}, B_{aba}, B_{baa}, B_{abc}, B_{acb}\}$			•		
$\{B_{aaa}, B_{aba}, B_{acb}\}$	•										

relations. In this section, we give a quick survey of the results of that kind, relating them to the concept of *conceptual space* introduced by Gärdenfors [9]. We then examine the question of interpreting the preceding results about points on a circle in that context.

4.1 Conceptual Spaces

Conceptual spaces are based on domains. A typical example of a domain is the *color domain* as represented by the Swedish natural color system (NCS) [11] which is a perceptual model of color perception. It describes the phenomenal structure of colors, that is, colors as we perceive them, using three dimensions: *hue*, *chromaticness* (or saturation), and *brightness*.

The first dimension, hue, is represented by the color circle. Colors lying opposite to each other are complementary colors: for example, green is complementary to red, orange to blue.

The second dimension, chromaticness, ranges from zero color intensity to increasingly greater intensities. It is modelled by a segment. Hence hue and chromaticness taken together are modelled by a disk, where colors can be distinguished on the periphery, and become more and more blurred as one comes closer to the center.

The third dimension is brightness which varies from white to black, and is consequently also represented by a segment. Brightness and chromaticness do not vary independently: variations in chromaticness decrease in range when brightness approaches black or white. Hence, for a given hue, the space of possible pairs (chromaticness, brightness) describes a triangle.

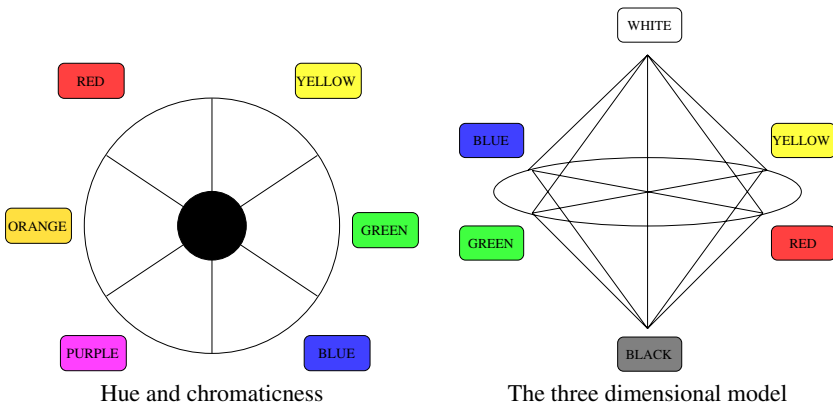


Fig. 2. The NCS model of colors.

Globally, then the model is called the NCS color spindle [27]. Gärdenfors gives a detailed discussion of the use of the model for explaining linguistic phenomena (such as the use of color terms), based on the assumption that terms referring to “natural” properties, that is in particular properties which can be named, correspond to *convex* subsets of the model.

The color model example is only a particular instance of the general hypothesis about natural properties: they should correspond to convex regions in some suitable conceptual model. The interested reader should refer to [9].

The NCS model is an example of a *phenomenal* conceptual space. Other conceptual spaces are theoretical conceptual spaces: For instance, the conceptual model of space in Newtonian physics is a 3-dimensional Euclidean space, time being an independent (in Gärdenfors’ terminology, separable) dimension. By contrast, the temporal dimension in relativistic physics is an integral dimension of the 4-dimensional Minkowski space.

4.2 The Conceptual Space of Allen’s Relations

As a first, and typical case of a conceptual space in the domain of qualitative temporal reasoning, we consider the case of the Allen calculus [1]. Since an interval on the real line is characterized by a (strictly) increasing pair of real numbers, a model of the set

of all intervals is the open half-plane delimited by the first bisector in the (X, Y) -plane. This half-plane is defined by the equation $Y > X$. Given a fixed interval (a, b) , with $a < b$, the basic Allen relations correspond to 13 regions in the half-plane, as shown in Fig. 3.

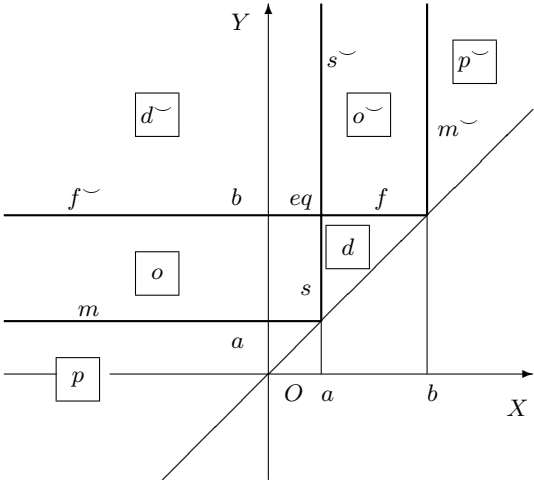


Fig. 3. The conceptual space of atomic relations in Allen’s algebra

The conceptual space of Allen’s relations has a much richer structure than the mere algebra. In particular, each relation has a dimension (the dimension of the corresponding region, which corresponds to the number of degrees of freedom of the relation). The incidence structure of the set of regions is a graph whose vertices correspond to the atomic relations, where there is an arc from r_1 to r_2 if r_2 belongs to the boundary of r_1 . This incidence structure can be deduced from the conceptual space, cf. Fig. 4. It contains enough topological information to encode the closure properties of the relations. In particular, the closure of any relation can be read from this graph.

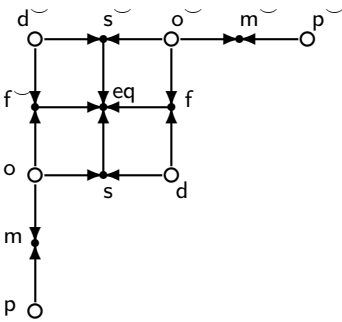


Fig. 4. The incidence graph of Allen’s relations

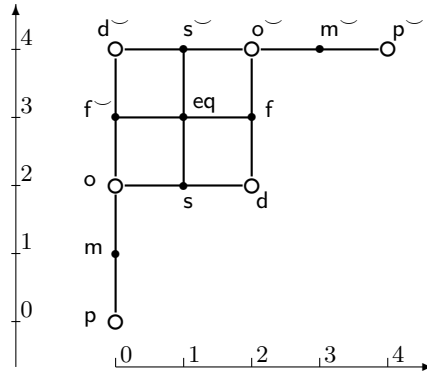


Fig. 5. The lattice of Allen's relations

Closely related to the incidence structure is the lattice of atomic relations represented in Fig. 5, which summarizes the order properties of the relations.

A basic problem in studying the complexity of reasoning with Allen's relations is the problem of determining whether a given constraint network is consistent. The general class of networks using any disjunction of atomic relations is known to be NP-complete. It is a remarkable fact that tractable subclasses of relations can be characterized in geometrical terms, as shown in [17] and subsequent papers [18,20].

Basic relations (as regions in the half-plane) are convex relations. In fact, they have a stronger property: they are also *saturated* (with respect to projections on the axes), in the sense that for such a region R , $R = pr_1(R) \times pr_2(R)$, where pr_1 and pr_2 are the X and Y projections respectively.

Convex relations are those unions of atomic relations which are both convex and saturated. In the lattice representation, this is equivalent to relations which are intervals in the lattice. More generally, *pre-convex* relations are those relations whose topological closure is a convex relation. Although they are neither convex nor saturated in general, they differ from the smallest convex closure by only "small" pieces, in the sense that the difference contains only relations whose dimension is strictly smaller. An argument based on this fact, together with the known fact that convex relations are tractable implies that the class of pre-convex relations is tractable [18]. In fact, it is the unique maximal tractable subclass containing all atomic relations [19].

Those results also can be obtained with purely syntactic methods: pre-convex relations coincide with ORD-Horn relations in the sense of [23], and their tractability is a consequence of the properties of Horn theories.

4.3 Conceptual Spaces and Complexity in Calculi Based on Linear Orderings

Allen's calculus fits into a larger family of calculi based on linear orderings:

- Generalized interval calculi [15,16,20] which consider finite strictly increasing sequences of points (in this context, Allen's calculus is the particular case of two points);

- The n -point calculus [4] where the basic objects are points in a n -D Euclidean space (the time point calculus is the case where $n = 1$). The case where $n = 2$ has been considered in [20] under the name of Cardinal Direction Calculus.
- The n -block calculus [5,2], whose basic objects are blocks (products of intervals) in a n -D Euclidean space (Allen's calculus is the case where $n = 1$).

Conceptual spaces are easily derived using the same method as for Allen's calculus in each particular case:

- In the case of generalized interval calculi, the conceptual space associated to (p, q) -relations, that is, relations from one p -interval to one q -interval ($p, q \geq 1$) is defined as follows:

Consider in the Euclidean q -space the cone C_q defined by:

$$X_1 < X_2 < \dots < X_q$$

Fix a point $(a_1, \dots, a_q)$ in C_q

The set of atomic (p, q) -relations $\Pi_{p,q}$ is the set of non-decreasing sequences of length p of integers between 0 and $2q$, where no odd integer occurs more than once. Each (p, q) -relation is associated with a region in C_q , and, globally, these regions constitute a partition of C_q .

The explicit definition of the region associated to a given (p, q) -relation $(b_1, \dots, b_p)$ is intricate but straightforward: if $b_i = 2n_i + 1$, then we consider the equation $X_i = a_{n_i}$; if $b_i = 2n_i$, then we consider the inequations $a_{n_i} < X_i < a_{n_i+1}$. The region is defined by the conjunctions of equations and inequations associated to each b_i , for $1 \leq i \leq p$.

Clearly, since it is defined by a Cartesian product of points or open intervals, (the region associated to) each atomic relation is convex and saturated.

Quite analogously to the case of Allen's relations, *convex* relations can be defined, e.g. as intervals in the lattice $\Pi_{p,q}$. They are convex and saturated. *Pre-convex* relations are those relations which have a convex topological closure. The same kinds of considerations as in Allen's case show that pre-convex relations, or more precisely the subclass of pre-convex relations, called strongly pre-convex relations [6] is tractable.

- For the n -point calculus, the conceptual space associated to basic relations of the n -point calculus (which are sequences of length n of point relations) is a partition of the Euclidean n -space:

Consider any point $a_1, \dots, a_n$ in the n -space.

A basic relation is a n -tuple $b_1, \dots, b_n$, where $b_i \in \{<, =, >\}$.

The region associated to such a basic relation is defined by the conjunction of the equations or inequations:

$X_i = a_i$ if b_i is $=$, $X_i < a_i$ if b_i is $<$, and $X_i > a_i$ if b_i is $>$.

Clearly again, these regions are convex and saturated. Convex, pre-convex, and strongly pre-convex relations can be defined, and tractability results obtained for strongly pre-convex relations [4].

- Finally, for the n -block calculus, the basic relations are sequences of Allen's relations. The corresponding conceptual space is a product of copies of the space for Allen's relations, and again, similar results obtain for pre-convex relations.

It must be mentioned, moreover, that in all three classes of calculi, strongly pre-convex relations coincide with ORD-Horn relations, which gives an independent motivation for their tractability, and constitutes a nice point of agreement between geometrically and syntactically motivated notions [6].

4.4 Points on a Circle

For all calculi based on linear orderings by taking sequences or products, the geometric structure of the basic relations, as represented in the corresponding conceptual space and in the lattice and incidence graph representations, are closely related to tractability properties. In line with the general considerations in Gärdenfors' framework, convexity and the stronger property of convexity plus saturation, play a crucial role.

The sad fact is that this does not seem to be the case any longer if we consider relations in the cyclic case. For the ternary relations between points considered in this paper, the incidence graph of the basic relations is easily obtained: starting from the relation where all three points coincide, that is, relation B_{aaa} , one gets, by separating either x , y , or z , one of the three relations B_{baa} , B_{aba} , B_{aab} . Going further and separating the remaining two points leads either to B_{abc} or to B_{acb} . Hence we get the graph in Fig. 4.4.

However, it is not at all clear how the (partial) complexity results we have obtained in this paper relate to geometric properties of this graph. This negative phenomenon may be related to the fact that, for the *binary* qualitative relations between intervals on a circle, path-consistent atomic networks may be inconsistent [7], or, in other terms, that weak-representations in the sense of [15,21] may well be inconsistent.

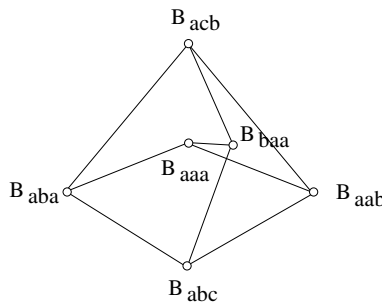


Fig. 6. The incidence graph of the cyclic point relations.

5 Conclusion

In a first part we described the relations between linear and cyclic models. Then, we considered six possible ternary relations between three points on a circle, which are jointly exhaustive and pairwise disjoint (JEPD) relations, and developed a qualitative calculus called the cyclic point algebra. We examined the consistency problem of the cyclic point networks. We have characterized several tractable and untractable cases for this problem.

The continuation of this work will be the complete characterization of all the tractable cases in the cyclic point algebra. Because of the small size of the set of relations of the cyclic point algebra, this goal seems to be reasonable.

Another perspective will be considering cyclic arcs instead of cyclic points. The relations considered will be those characterized by Balbiani and Osmani in [7]. A first task will consist in defining an axiom system for these relations. To this end, we can use the axiom system of the cyclic orders (see [13] for a similar work). Concerning the constraint aspects, our study of cyclic point networks can be certainly used to characterize new tractable cases for the consistency problem of the cyclic arc networks.

We presented the basic notions of the framework of conceptual spaces, in relation to the characterization of tractable subclasses for formalisms such as the Interval Algebra. An open question is: is there a geometric and topological characterization of tractable classes in the cyclic cases? It appears that finding a suitable conceptual space is more difficult (less natural) than in the linear case.

References

1. J. F. Allen. Maintaining knowledge about temporal intervals. *Comm. of the ACM*, 26(11):832–843, 1983.
2. Ph. Balbiani, J.-F. Condotta, and L. Fariñas del Cerro. A model for reasoning about bidimensional temporal relations. In *Proc. of KR-98*, pages 124–130, 1998.
3. Ph. Balbiani, J.-F. Condotta, and L. Fariñas del Cerro. A new tractable subclass of the rectangle algebra. In *Proc. of IJCAI-99*, pages 442–447, 1999.
4. Ph. Balbiani, J.-F. Condotta, and L. Fariñas del Cerro. Spatial reasoning about points in a multidimensional setting. In *Proc. of the IJCAI-99 Workshop on Spatial and Temporal Reasoning*, pages 105–113, 1999.
5. Ph. Balbiani, J.-F. Condotta, and L. Fariñas del Cerro. A tractable subclass of the block algebra: constraint propagation and preconvex relations. In *Proc. of the Ninth Portuguese Conference on Artificial Intelligence (EPIA'99)*, pages 75–89, 1999.
6. Ph. Balbiani, J.-F. Condotta, and G. Ligozat. Reasoning about Generalized Intervals: Horn Representation and Tractability. In Scott Goodwin and Andre Trudel, editors, *Proceedings of the Seventh International Workshop on Temporal Representation and Reasoning (TIME-00)*, pages 23–30, Cape Breton, Nova Scotia, Canada, 2000. IEEE Computer Society.
7. Ph. Balbiani and A. Osmani. A model for reasoning about topologic relations between cyclic intervals. In *Proc. of KR-2000*, Breckenridge, Colorado, 2000.
8. A. U. Frank. Qualitative spatial reasoning about distances and directions in geographic space. *J. of Visual Languages and Computing*, 3:343–371, 1992.
9. P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. The MIT Press, 2000.
10. M. R. Garey and D. S. Johnson. *Computers and intractability; a guide to the theory of NP-completeness*. W.H. Freeman, 1979.

11. A. Hård and L. Sivik. NCS-natural color system: a Swedish standard for color notation. *Color Research and Application*, (6):129–138, 1981.
12. A. Isli and A. G. Cohn. A new approach to cyclic ordering of 2D orientations using ternary relation algebras. *Artificial Intelligence*, 122(1–2):137–187, 2000.
13. P. Ladkin. *The Logic of time representation*. PhD thesis, University of California, Berkeley, 1987.
14. C. Langford. Some theorems on deducibility. *Ann. Math. Ser.*, 28:16–40, 1927.
15. G. Ligozat. Weak Representations of Interval Algebras. In *Proc. of AAAI-90*, pages 715–720, 1990.
16. G. Ligozat. On generalized interval calculi. In *Proc. of AAAI-91*, pages 234–240, 1991.
17. G. Ligozat. Tractable relations in temporal reasoning: pre-convex relations. In F. D. Anger, H. Güsgen, and G. Ligozat, editors, *Proc. of the ECAI-94 Workshop on Spatial and Temporal Reasoning*, pages 99–108, Amsterdam, 1994.
18. G. Ligozat. A New Proof of Tractability for ORD-Horn Relations. In *Proc. of AAAI-96*, pages 395–401, 1996.
19. G. Ligozat. "Corner" relations in Allen's algebra. *CONSTRAINTS: An International Journal*, 3:165–177, 1998.
20. G. Ligozat. Reasoning about Cardinal Directions. *J. of Visual Languages and Computing*, 9:23–44, 1998.
21. G. Ligozat. Simple Models for Simple Calculi. In C. Freksa and D.M. Mark, editors, *Proc. of COSIT'99*, number 1661 in LNCS, pages 173–188. Springer Verlag, 1999.
22. R. Moratz, J. Renz, and D. Wolter. Qualitative spatial reasoning about line segments. In Horn. W., editor, *ECAI 2000. Proceedings of the 14th European Conference on Artificial Intelligence*, Amsterdam, 2000. IOS Press.
23. B. Nebel and H.-J. Bürckert. Reasoning about temporal relations: A maximal tractable subclass of Allen's interval algebra. *J. of the ACM*, 42(1):43–66, 1995.
24. R. Röhrig. Representation and processing of qualitative orientation knowledge. In Gerhard Brewka, Christopher Habel, and Bernhard Nebel, editors, *Proceedings of the 21st Annual German Conference on Artificial Intelligence (KI-97): Advances in Artificial Intelligence*, volume 1303 of *LNAI*, pages 219–230, Berlin, September 9–12 1997. Springer.
25. C. Schlieder. Representing visible locations for qualitative navigation. In N. Piera Carreté and M. G. Singh, editors, *Proceedings of the III IMACS International Workshop on Qualitative Reasoning and Decision Technologies—QUARDET'93—*, pages 523–532, Barcelona, June 1993. CIMNE.
26. C. Schlieder. Reasoning about ordering. In *Proc. of COSIT'95*, 1995.
27. L. Sivik and C. Taft. Color naming: a mapping in the NCS of common color terms. *Scandinavian Journal of Psychology*, (35):144–164, 1994.
28. T. Sogo, H. Ishiguro, and T. Ishida. Acquisition of qualitative spatial representation by visual observation. In *Proceedings IJCAI-99*, pages 1054 – 1060, 1999.

Reasoning and the *Visual-Impedance* Hypothesis

Markus Knauff¹ and P.N. Johnson-Laird²

¹Freiburg University, Center for Cognitive Science
Friedrichstr. 50, D-79098 Freiburg, Germany
knauff@cognition.iig.uni-freiburg.de

²Princeton University, Department of Psychology
Green Hall, Princeton, NJ 08544, USA
phil@princeton.edu

Abstract. The *visual-impedance* hypothesis postulates that relational expressions which elicit visual images without a spatial component impede reasoning (Knauff and Johnson-Laird, in press). The goal of the present article is to summarize some experimental findings that support this hypothesis. Previous studies yielded four sorts of relations: (1) visuo-spatial relations, such as "above-below", that are easy to envisage visually and spatially, (2) visual relations, such as "cleaner-dirtier" that are easy to envisage visually but hard to envisage spatially, (3) spatial relations, such as "ancestor of-descendant of", that are hard to envisage visually but easy to envisage spatially and (4) control relations, such as "better-worse", that are hard to envisage either visually or spatially. Two behavioral studies showed that visual relations slow down reasoning in comparison with control relations, whereas visuo-spatial and spatial relations yield inferences comparable to those of control relations. The results of an fMRI study showed that in the absence of any correlated visual input (problems were presented acoustically via headphones) reasoning about all four sorts of relations evoked activity in the left middle temporal gyrus, in the right superior parietal cortex, and bilaterally in the precuneus. However, only the visual relations also activated areas of the primary visual cortex corresponding to Brodmann's area 18 (V2). The findings corroborate the theory that individuals rely on mental models for deductive reasoning, and that visual imagery irrelevant to reasoning impedes the process.

1 Introduction

Images are an important part of human cognition and it is natural to suppose that they can help humans to reason. This view is supported by various sorts of evidence including the well-known studies of the mental rotation and the mental scanning of images (Shepard and Cooper, 1982; Kosslyn, 1980). Moreover, several studies have shown that reasoning depends on the ease of imagining the premises, the instructions to form images, and the participants' ability to form images (e.g., Shaver, Pierson, and Lang, 1976; Clement and Falmagne, 1986). In contrast, however, other studies have failed to detect any effect of imageability on reasoning. Sternberg (1980) found no difference between the accuracy of solving problems that were easy or hard to visualize. Richardson (1987) reported that reasoning with visually concrete problems was no better than reasoning with abstract problems. Johnson-Laird, Byrne and Tabossi

(1989) examined reasoning with three transitive relations that differed in imageability: *equal in height*, *in the same place as*, and *related to* (in the sense of kinship). They did not find any effect of imageability on reasoning accuracy. Newstead, Pollard, and Griggs (1986) had reported similar results. In Knauff (2001), Knauff and Johnson-Laird (2000), and Knauff and Johnson-Laird (in press) we postulated that a possible resolution of the inconsistency in the results is that investigators have overlooked the distinction between *visual images* and *spatial representations*. We formulated the following *hypothesis*:

Visual-impedance hypothesis: Relations that elicit visual images without a component relevant to inference impede the process of reasoning.

The distinction between visual and spatial processes was originally detected in lesion studies with monkeys (Ungerleider and Mishkin, 1982) and in experiments with humans with brain injuries (for a review, see Newcombe and Ratcliff, 1989). These studies showed that visual and spatial processes are associated with different cortical areas. Additional support for the distinction comes from experiments examining human working memory (e.g. Logie, 1995) and most recently from functional brain imaging studies (e.g. D'Esposito, 1998; Smith, et al, 1995).

But, what does the distinction between visual and spatial imagery mean for reasoning? And how can the *visual-impedance* hypothesis be justified? On the one hand, a relation such as: *The hat is above the cup*, is easy to visualize given a modicum of competence in forming images. However, it can also be readily represented spatially. That is, individuals can construct a spatial model of the relation without any conscious awareness of a visual image. According to the theory of mental models, such a model suffices for reasoning. It captures the relevant logical properties. Hence, the transitivity of a relation of the form: *A is above B*, derives merely from the meaning of the relation and its contribution to models of assertions. Given premises of the form:

- A is above B.
- B is above C.

Reasoners build a two- or three-dimensional mental model that satisfies the premises:

A
B
C

This model supports the conclusion: *A is above C*, and no model of the premises refutes this conclusion (see Johnson-Laird and Byrne, 1991). Mental models are therefore not to be identified with visual images. Models are abstract, but they make it possible in certain cases to construct a visual image from a particular point of view (see e.g. Johnson-Laird, 1998).

On the other hand, a relation such as: *The hat is dirtier than the cup*, is easy to visualize, but it seems much less likely to be represented spatially. Subjectively, one seems to form an image of a dirty hat and an image of a less dirty cup. Such an image contains a large amount of information that is irrelevant to the inference, and so it puts an unnecessary load on working memory. In addition, reasoners have to isolate the information that is relevant to the inference. And so they might be side-tracked by

the irrelevant visual details. A visual image of, say, a dirty hat and a dirty cup gets in the way of forming a representation that makes the transitive inference possible.

In the next section, we summarize two behavioral experiments that test the *visual-impedance* hypothesis. We then present some results from an fMRI study on imageability and reasoning. For the benefit of the interdisciplinary readership of the present book, we refrain from reporting experimental details, and we discuss only those results that are statistically reliable (for further details, see Knauff, Fangmeier, Ruff, and Johnson-Laird, 2002 and Knauff and Johnson-Laird, in press).

2 Behavioral Experiments

The aim of the experiments is was to test the *visual-impedance* hypothesis. In Experiment 1, we examined reasoning with three sorts of relations:

1. Visuo-spatial relations that are easy to envisage visually and spatially (*above* and *below*, *in front of* and *to the back of*).
2. Visual relations that are easy to envisage visually but hard to envisage spatially (*cleaner* and *dirtier*, *fatter* and *thinner*).
3. Control relations that are hard to envisage both visually and spatially (*better* and *worse*, *smarter* and *dumber*).

The relations were selected from those in a study in which students from Princeton University rated the ease of envisaging a set of relations as visual images and as spatial layouts (Knauff and Johnson-Laird, 2000). In this study, we had examined the three types of relations in transitive inferences (Knauff and Johnson-Laird, 2000). However, it is possible that such inferences favor certain reasoning strategies. Hence, in the present experiment we examined the three sorts of relations in reasoning that combined conditional and relational reasoning. If our *visual-impedance* hypothesis is correct, then visual relations will slow down reasoning in comparison with the visuo-spatial and control relations. But, if the orthodox imagery hypothesis is correct, then participants should perform better with visual relations. We also manipulated the difficulty of the inferences.

The participants had to evaluate conditional inferences in the form of modus ponens, e.g.:

If the ape is smarter than the cat, then the cat is smarter than the dog.

The ape is smarter than the cat.

Does it follow:

The ape is smarter than the dog?

All the inferences used the same nouns (*dog*, *cat*, *ape*) in order to minimize differences as a result of anything other than the relations. The difficulty of the inferences was manipulated by using converse relations. The easiest inferences were of the form exemplified in the preceding example:

1. If aRb then bRc
aRb
aRc?

where aRb denotes a proposition asserting that a transitive relation, R , holds between two entities, a and b . The converse relation, R' , such as *dumber*, yields more difficult inferences in the following forms:

2. If aRb then $cR'b$ aRb $aRc?$	3. If $bR'a$ then bRc $bR'a$ $aRc?$

Reasoners now have to convert $cR'b$ in order to make the transitive inference. The hardest form of inference used two separate converse relations, one in the premise and one in the conclusion:

4. If $bR'a$ then bRc
 $bR'a$
 $cR'a?$

The participants acted as their own controls and evaluated two valid and two invalid inferences at the three levels of difficulty for each of the three sorts of relations (visual, visuo-spatial, control), making a total of 36 problems.

Overall, the participants made 71% correct responses. The trend concerning the effect of the converse relations on the percentages of correct responses fell short of significance: The easiest problems (Type 1) yielded 86% correct responses with a mean latency of 2.6s, the intermediate problems with one converse relation (Types 2 and 3) yielded 73% correct responses with a mean latency of 2.7s, and the hardest problems (Type 4) yielded 54% correct problems with a mean latency of 3.3s (Page's $L = 243$, $p > .05$, and $L = 242$, $p > .05$, respectively). There was no significant difference in accuracy for the three sorts of relations at any level of difficulty, and there was no significant interaction between the two variables. Likewise, the three sorts of relations did not have a significant effect on accuracy at any level of difficulty, and there was no significant interaction between the two variables (Wilcoxon test, $z = 0.58$, $p > .56$).

In contrast to the results on accuracy, the latencies of the responses corroborated the *visual-impedance* hypothesis. Figure 1 presents the mean latencies for the correct responses to the inferences based on the three sorts of relations. There was a reliable trend: Responses were faster to the visuo-spatial inferences (2456 ms) than to the control inferences (2643 ms), which in turn were faster than the visual inferences (3365 ms; Page's $L = 255$, $p < .05$). The difference between the visuo-spatial inferences and the control inferences was not significant, but the control inferences were reliably faster than the visual inferences (Wilcoxon test $z = 2.46$; $p < .02$). There was no reliable interaction between the relations and the levels of difficulty in their effects on latencies (Wilcoxon test, $z = 0.75$, $p > .45$).

These results show that the visual relations slowed down reasoning in comparison with control relations, which were harder to visualize. There was a tendency, though it was not significant, for the visuo-spatial relations to yield slightly faster responses than the control relations.

What happens if a relation is easy to envisage spatially but not easy to visualize? Our previous rating studies failed to discover any such relations. But, if materials that are easy to visualize impair reasoning, whereas materials that are easy to envisage spatially speed up reasoning, then reasoning based on purely spatial relations should be the fastest.

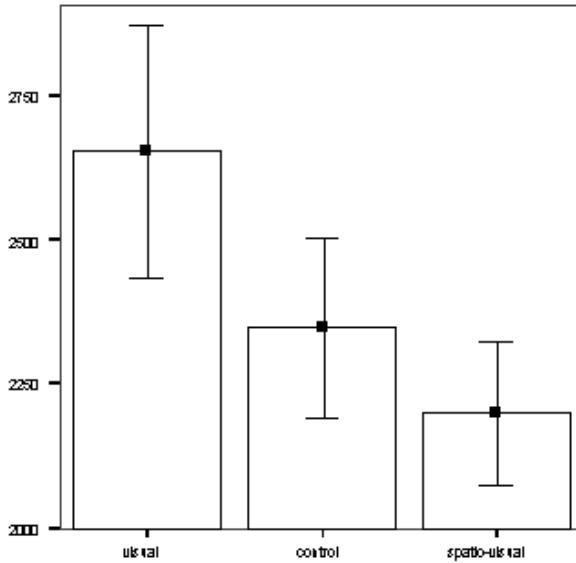


Fig. 1. Mean response latencies [in ms] and standard errors in reasoning in Experiment 1 with three sorts of relations: visual relations, control relations and visuo-spatial relations.

We therefore renewed our search for such relations. We asked twelve native German speakers at Freiburg University to complete a questionnaire in which they rated the ease of forming visual images and spatial layouts for a set of relational assertions. We included the relations from the first rating study but added more relations, which seemed easy to envisage spatially but difficult to visualize (*earlier* and *later*, *older* and *younger*, *hotter* and *colder*, *faster* and *slower*, *further North* and *further South*, *stronger* and *weaker*, *bigger* and *smaller*, *ancestor of* and *descendant of*, and *heavier* and *lighter*). The results replicated those from the earlier study at Princeton University and yielded the same three sorts of relations. We concluded that the procedure of separate ratings might not be sensitive enough to reveal purely spatial relations. We therefore carried out a study using a different procedure.

The participants rated each relation on a single bipolar seven-point scale, ranging from ease of evoking a "visual" image at one end of the scale to ease of evoking a "spatial" layout at the other end of the scale. The instructions stated that a visual representation is a vivid visual image that can include people, objects, colors, and shapes, and that it can be similar to a real perception. They stated that a spatial representation is a more abstract layout and represents something on a scale or axis, or in a spatial array. We tested 20 students with a set of 35 relations.

The results revealed two pairs of purely spatial relations: *ancestor of* and *descendant of*, and *further North* and *further South* (in German, *nördlicher* and *südlicher*, which are single words). The ratings for the four sorts of relations differed significantly (Friedman analysis of variance $F = 38.33$; $p < .001$). With these relations, we carried out a second experiment.

Experiment 2 examined reasoning with the four sorts of relations (visual, spatial, visuo-spatial, and controls). The *visual-impedance* hypothesis predicts that the visual relations should slow down reasoning. If the construction of a spatial representation

speeds up reasoning, even in the absence of visualization, then both the spatial and the visuo-spatial relations should speed up reasoning in comparison with the control relations. Hence, the four relations should show the following trend in increasing latencies for reading and reasoning: spatial, visuo-spatial, control, and visual.

The materials consisted of 16 three-term and 16 four-term series inferences. All the inferences again used the same nouns (*dog*, *cat*, *ape*, and for four-term inferences: *bird*). Here is an example of a three-term inference with a valid conclusion:

The dog is cleaner than the cat.
 The ape is dirtier than the cat.
 Does it follow that:
 The dog is cleaner than the ape?

And here is an example of a four-term series inference with an invalid conclusion:

The ape is smarter than the cat.
 The cat is smarter than the dog.
 The bird is dumber than the dog.
 Does it follow that:
 The bird is smarter than the ape?

There were two valid and two invalid inferences using each of the four sorts of relations in both three-term and four-term series inferences, making a total of 32 inferences. The 24 participants acted as their own controls and evaluated the 32 inferences presented in random order.

Overall, the participants responded correctly to 74% of the inferences and there was no significant difference in error rates for the different sorts of inferences. The mean latencies for the correct responses to the four sorts of relations are shown in Figure 2. The fastest response was for the spatial relations (3516ms), followed by the visuo-spatial relations (3736ms), the control relations (3814ms), and the visual relations (4482ms). This trend was statistically significant (Page's $L = 648$, $z = 3.40$, $p < .05$). However, as Figure 2 suggests, the only significant effect is that visual relations slow down reasoning to a greater extent than the other three relations (Wilcoxon test $z = 2.46$; $p < .015$).

The first experiment showed that visual relations significantly impeded the process of reasoning, whereas visuo-spatial relations yielded response latencies comparable to those of control relations. The second experiment showed that purely spatial relations, which are difficult to envisage visually but easy to envisage spatially, yield slightly faster inferences, though the trend was not reliable. In both experiments, however, visual relations impeded reasoning.

Some accounts of reasoning postulate that inferences are based on visual images, which are similar in structure to actual percepts, and which can represent colors, shapes, and spatial extent. Images can be rotated and scanned, and they have a limited resolution (Kosslyn, 1980). Mental operations on an image can be isomorphic to those on real percepts. Similarly, an image can be confused in memory with a real percept (Johnson and Raye, 1981). Reasoning based on images calls for individuals to "look" at the image based on the premises, and to "read off" a conclusion not explicitly stated in the premises. This account of reasoning has difficulty in explaining our results. If

reasoning is based on visual images, then it is hard to understand why the visual relations in our studies slowed down reasoning performance.

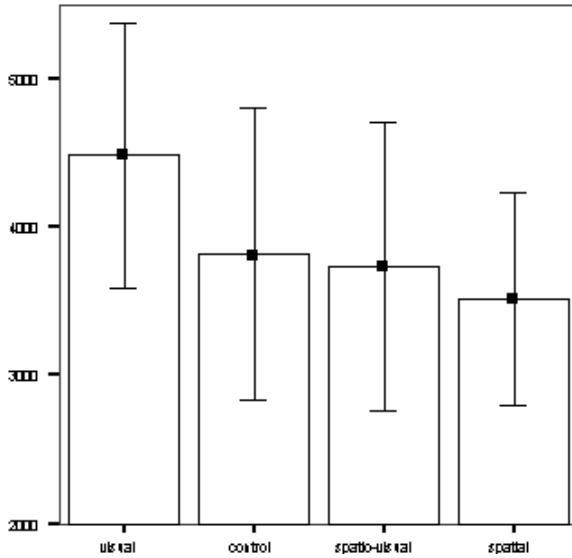


Fig. 2. Mean response latencies [in ms] and standard errors in relational reasoning with four sorts of relations: visual relations, control relations, visuo-spatial relations, and spatial relations.

What our results suggest is that in many cases reasoning is based not on visual images, but on more abstract structures, i.e., mental models. These representations avoid excessive visual detail in order to bring out salient information for inferences (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991).

3 A Functional Brain Imaging Experiment

We have performed several recent studies of reasoning and visual imagery using functional magnetic resonance imaging (fMRI). In a study by Knauff, Mulack, Kas-subek, Salih, and Greenlee (2002), for instance, conditional and relational reasoning activated a bilateral parietal-frontal network distributed over parts of the prefrontal cortex, the inferior and superior parietal cortex, and the precuneus, whereas no significant activation occurred in the occipital cortex, which is usually activated by visual imagery (Kosslyn et al., 1993; Kosslyn et al., 1999; Kosslyn, Thompson, and Alpert, 1997; Sabbah et al., 1995; a contrasting result is reported in Knauff, Kas-subek, Mulack, and Greenlee, 2000). In fact, reasoning activated regions of the brain that make up the “where-pathway” of spatial perception and working memory (e.g., Ungerleider and Mishkin, 1982; Smith et al., 1995). In contrast, the “what-pathway” that processes visual features such as shape, texture, and color (cf. also Landau and Jackendoff, 1993; Rueckl, Cave, and Kosslyn, 1989; Ungerleider, 1996) seemed not to be activated. Other experiments have corroborated these findings. Prabhakaran, Smith, Desmond, Glover, and Gabrieli (1997) studied Raven's Progressive Matrices

and found (for inductive reasoning) increased activity in right frontal and bilateral parietal regions. Osherson et al. (1998) compared inductive and deductive reasoning and found that the latter increased activation in right-hemisphere parietal regions. Goel and Dolan (2001) studied concrete and abstract three-term relational reasoning and found activation in a parietal-occipital-frontal network. Kroger, Cohen, and Johnson-Laird (2001) found that reasoning in contrast to mental arithmetic based on the same assertions activated right frontal areas often associated with spatial representation.

What happens in the brain if participants solve problems with the four sorts of relations from the behavioral experiments (visual, visuo-spatial, spatial, and control)? Are the differences in imageability reflected in differences in brain activation? The behavioral experiments showed that reasoning with visual relations was more difficult than with the other relations, but there was no significant difference between visuo-spatial, spatial, and control problems. The *visual-impedance* effect appears to occur because visual details are irrelevant to inference, and it takes additional time to retrieve the relevant information. To test whether visual relations do indeed elicit visual images, we carried out a brain-imaging experiment (Knauff, Fangmeier, and Ruff, 2002; Knauff, Fangmeier, Ruff, and Johnson-Laird, 2002).

Experiment 3 examined the four sorts of relations. The participants were 12 healthy male right-handed volunteers. The reasoning problems were identical to the transitive inferences used in Experiment 2. The participants' task was to decide whether or not a given conclusion followed from the premises. They made their response by pressing the appropriate key. The problems were presented verbally via pneumatic headphones, eliminating the need for visual input. There were eight problems for each of the four sorts of relations, yielding a total of 32 problems. The 32 problems were presented in four separate runs, each contained four blocks with one problem pair for each of the problem types (visuo-spatial, visual, spatial, and control). The problem pairs were randomly determined for each problem type and they remained constant throughout the experiment for all participants. The problems were randomly assigned to the runs of each subject. The order of the problem within one run was also randomly determined. Half of the problems were valid, the other half invalid. The inference tasks were identical in all respects, except for the nature of the relations. A rest interval of similar length was included between problems which differed only in lack of problem presentation. The details can be found in Knauff, Fangmeier, Ruff, and Johnson-Laird (2002).

The response latencies showed a similar pattern to those of Experiment 2; correct responses were slower for the visual problems (2.1 s) than for the control inferences (2.0 s), visuo-spatial (2.0 s), and spatial inferences (2.0 s). The differences were not reliable, however, probably because of the small sample size (Friedman analysis of variance $F = 4.64$; $p = 0.20$).

Although the control problems were the baseline condition for assessing differences in the neural processing of the different sorts of relations, the additional rest condition was initially used to determine the activation evoked by the entire set of reasoning problems. Hence, the analysis of the imaging data was carried out in two steps. The first analysis was performed to identify the cortical areas active for reasoning in general (visual, visuo-spatial, spatial, and control problems vs. the rest condition). The second analysis was carried out to examine differences among the four sorts of relation. We expected that reasoning in general should evoke activity in the spatial (dorsal) pathway, in particular in BA 7. But, if the participants generated vis-

ual images for the visual relations, then only these relations should activate areas of the brain devoted to the processing of visual information.

The results corroborated our predictions. Reasoning in general led to bilateral activity in parietal cortices. The first analysis showed that activation was similar for the four sorts of relation in comparison with the rest period. The active parietal areas in this contrast are presented in Figure 3. The figure shows that all four sorts of reasoning led to bilateral activity in the precuneus (BA 7), and in right superior parietal cortex (BA 40).

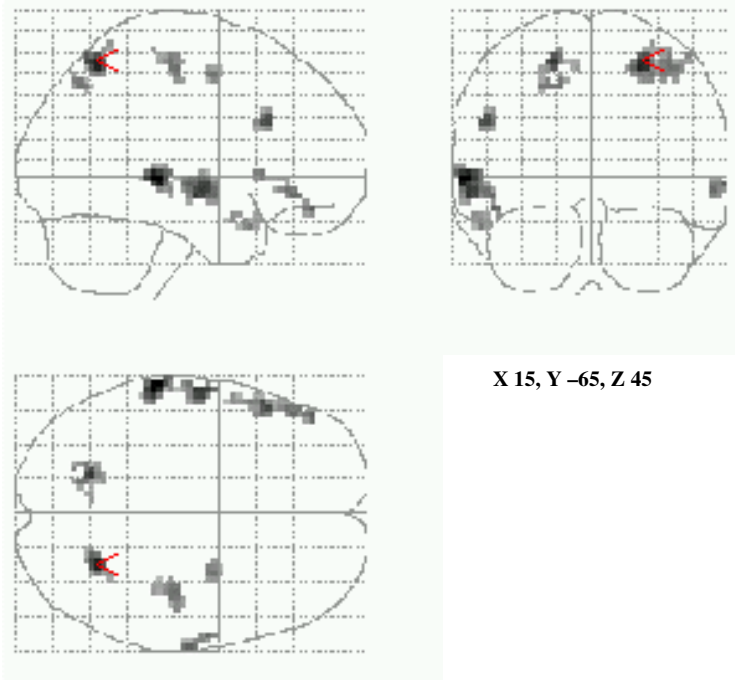


Fig. 3. All four sorts of reasoning activated the bilateral parietal cortex. The figure shows the contrast in activation between the four sorts of reasoning problem and the rest condition. In the pictures, all activities were transferred to an arbitrary gray scale, and projected onto sagittal, coronal, and transverse sections of a standard brain template. Slice positions according to the Talairach atlas are given in the lower right corner of the figure (X, Y, Z, coordinates). Cross-hairs are positioned in the local peak voxel for the respective contrast and brain area.

The second analysis compared reasoning with the control relations with each of the other sorts of relation: visuo-spatial, visual, and spatial. It showed that only the visual relations led to additional activation in an area that covers parts of the visual association cortex (corresponding to BA 18) and the precuneus (BA 31). These additional areas are shown in Figure 4.

Experiment 3 and previous studies (Goal and Dolan, 2001; Knauff, Mulack, Kas-subek, Salih, and Greenlee, 2002) yield a consistent pattern of results: a neural correlate of deductive reasoning is located in a bilateral occipito-parietal-frontal network distributed over parts of the prefrontal cortex and the cingulate gyrus, the superior parietal cortex, and the precuneus. The parietal cortex is considered to be an area that

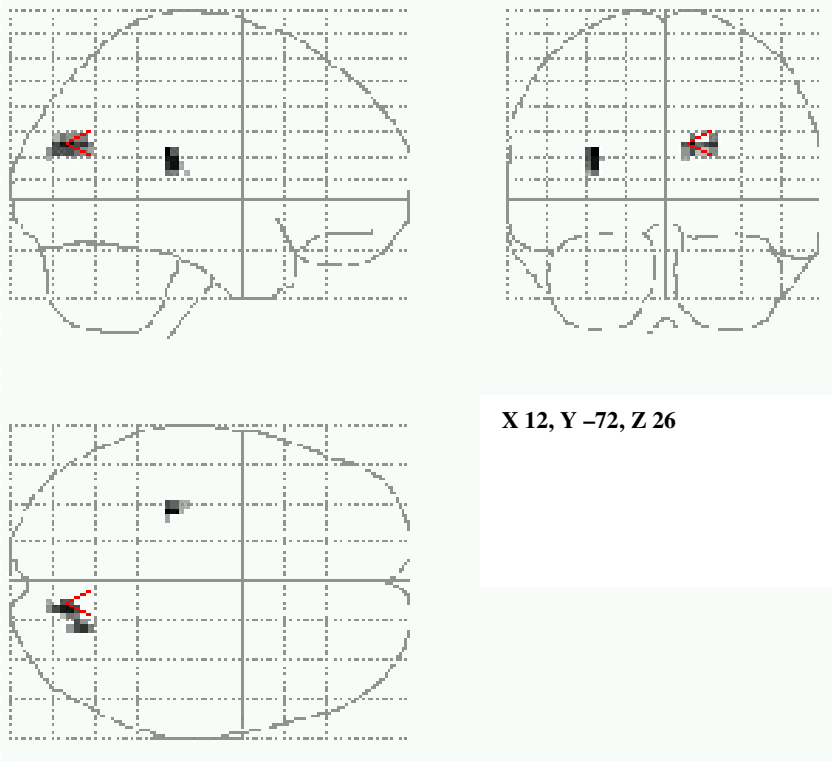


Fig. 4. Cortical regions significantly activated by reasoning with visual relations, such as *dirtier* and *cleaner* as compared to control relations). The figure shows the activity in the occipital cortex, corresponding to secondary visual cortex (V2, BA 18). Slice positions according to the Talairach atlas are given in the lower right corner of the figure. Crosshairs are positioned in the local peak voxel for the respective contrast and brain area.

combines information from different sensory modalities to form cognitive representations of space (see, e.g., Andersen, 1997). Hence, the results suggest that deductive reasoning is based on spatial representations and processes. In addition, the visual relations activated regions in visual association cortex (V2). This result corroborates the *visual-impedance* hypothesis. Visual details impede reasoning because they are irrelevant to the process. The hypothesis is accordingly borne out by the additional activity in visual areas of the brain. This activity occurred only with the visual relations, and the activated areas were not in primary visual cortex, but in visual association areas. Thus, the strongest hypothesis that imagery evokes activity in primary visual cortex (e.g. Kosslyn, 1994) was not supported by the present data. But, several other imaging studies have shown that mental imagery does not necessarily activate the primary visual cortex, but higher visual areas (e.g. Knauff, Kassubek, Mulack, and Greenlee, 2000).

4 Conclusions

The starting point of our experiments was the assumption that the conflicting results in the literature on mental imagery and deductive reasoning arose from a failure to distinguish between visual and spatial modes of representation. We accordingly proposed a *visual-impedance* hypothesis: relations that elicit visual images without a component relevant to inference impede reasoning. The behavioral experiments supported this hypothesis. Moreover, the impedance effect resolves some of apparent inconsistencies in the literature. Those studies that found a facilitating effect of imagery tended to use materials that differed in the ease of constructing spatial representations, whereas those studies that found no such effect, or an impeding effect of imageability, tended to use materials that evoked visual representations (see Knauff and Johnson-Laird, in press).

The brain imaging experiment provided further evidence that visual impedance is a result of the spontaneous tendency to construct visual images when the material is easy to visualize. These visual images are usually irrelevant for reasoning. Our reasoning problems were so easy that such irrelevant visual images were unlikely to lead individuals into error; but they did slow down the process. The inferential system has to find the pertinent information amongst the details and may have to suppress the irrelevant visual detail. One corollary is that visual imagery is *not* a mere *epiphenomenon* playing no causal role in reasoning (e.g. Pylyshyn, 1981). It can even be a nuisance in thinking.

Acknowledgments

This research was supported in part by a grant from the German National Research Foundation (DFG) to the first author (WorkSpace, Grant Kn465/2-3) to study reasoning and working memory, and to the second author from the American National Science Foundation (NSF; Grant BCS 0076287) to study strategies in reasoning. The authors are grateful to Elin Arbin, Uri Hasson, Emily Janus, Juan Garcia Madruga, Thomas Fangmeier, Christian Ruff, Vladimir Sloutsky, Gerhard Strube, Clare Walsh, Yingrui Yang, and Lauren Ziskind, for helpful discussions of the research.

References

- Andersen, R. A. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical transactions of the Royal Society of London. Series B: Biological Sciences*, 352, 1421-1428.
- Clement, C., A. & Falmagne, R., J. (1986). Logical reasoning, world knowledge, and mental imagery: Interconnections in cognitive processes. *Memory & Cognition*, 14, 299-307.
- D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research*, 7, 1-13.

- Goel, V., & Dolan, R. J. (2001). Functional neuroanatomy of three-term relational reasoning. *Neuropsychologia*, 39, 901-909.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N. (1998). Imagery, visualization, and thinking. In J. Hochberg (Ed.), *Perception and Cognition at Century's End* (pp. 441-467). San Diego, CA: Academic Press.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates.
- Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, 96, 658-673.
- Knauff, M. (2001). *Vivid reasoning, mind, and brain*. Habilitationsschrift. Freiburg: Philosophische Fakultät I der Universität Freiburg.
- Knauff, M., Fangmeier, T., & Ruff, C. C. (2002). Vividness, mental imagery, and deductive reasoning: a study using functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, Supplement*, 68.
- Knauff, M., Fangmeier, T., Ruff, C., & Johnson-Laird, P.N. (2002). Reasoning, models, and images: Behavioral measures and cortical activity. *Under submission*.
- Knauff, M., & Johnson-Laird, P. N. (2000). Visual and spatial representations in spatial reasoning. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 759-765). Mahwah, NJ: Lawrence Erlbaum Associates.
- Knauff, M., & Johnson-Laird, P. N. (in press). Visual imagery can impede reasoning. *Memory & Cognition*.
- Knauff, M., Kassubek, J., Mulack, T., & Greenlee, M. W. (2000). Cortical activation evoked by visual mental imagery as measured by functional MRI. *NeuroReport*, 11, 3957-3962.
- Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., & Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: a functional MRI study. *Cognitive Brain Research*, 13, 203-212.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.
- Kroger, J., Cohen, J., and Johnson-Laird, P.N. (2001) A double dissociation between logic and mathematics: a functional magnetic resonance imaging study. *Under submission*.
- Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and brain sciences*, 16, 217-265.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hove: Lawrence Erlbaum Associates.
- Newcombe, F., & Ratcliff, G. (1989). Disorders of visuospatial analysis. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology* (Vol. 2, pp. 333-356). Amsterdam: Elsevier.
- Newstead, S. E., Pollard, P., & Griggs, R. A. (1986). Response bias in relational reasoning. *Bulletin of the Psychonomic Society*, 24, 95-98.

- Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive Psychology*, 33, 43-63.
- Pylyshyn, Z. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological review*, 88, 16-45.
- Richardson, J. T. E. (1987). The role of mental imagery in models of transitive inference. *British Journal of Psychology*, 78, 189-203.
- Rueckl, J. G., Cave, K. R., & Kosslyn, S. M. (1989). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience*, 1, 171-186.
- Shaver, P., Pierson, L., & Lang, S. (1974). Converging evidence for the functional significance of imagery in problem solving. *Cognition*, 3, 359-375.
- Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Sternberg, R. J. (1980). Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General*, 109, 119-159.
- Ungerleider, L. G. (1996). Functional brain imaging studies of cortical mechanisms for memory. *Science*, 270, 769-775.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of Visual Behaviour* (pp. 549-587). Cambridge, MA: MIT Press.

Qualitative Spatial Reasoning about Relative Position

The Tradeoff between Strong Formal Properties and Successful Reasoning about Route Graphs

Reinhard Moratz¹, Bernhard Nebel², and Christian Freksa¹

¹ Universität Bremen, Department of Mathematics and Informatics,
Bibliothekstr. 1, 28359 Bremen, Germany

{moratz|freksa}@informatik.uni-bremen.de

² University of Freiburg, Institute for Informatics
Georges-Köhler-Allee 52, D-79110 Freiburg, Germany
nebel@informatik.uni-freiburg.de

Abstract. Qualitative knowledge about relative orientation can be expressed in form of ternary point relations. In this paper we present a calculus based on ternary relations. It utilises finer distinctions than previously published calculi. It permits differentiations which are useful in realistic application scenarios that cannot directly be dealt with in coarser calculi. There is a price to pay for the advanced options: useful mathematical results for coarser calculi do not hold for the new calculus. This tradeoff is demonstrated by a direct comparison of the new calculus with the flip-flop calculus.

Keywords: Qualitative Spatial Reasoning, Cognitive Modelling, Robot Navigation

1 Introduction

Qualitative Spatial Reasoning (QSR) abstracts from metrical details of the physical world and enables computers to make predictions about spatial relations, even when precise quantitative information is not available [Cohn, 1997]. From a practical viewpoint QSR is an abstraction that summarizes similar quantitative states into one qualitative characterization. A complementary view from the cognitive perspective is that the qualitative method *compares* features within the object domain rather than by *measuring* them in terms of some artificial external scale [Freksa, 1992]. This is the reason why qualitative descriptions are quite natural for humans.

The two main directions in QSR are topological reasoning about regions [Randell et al., 1992], [Renz and Nebel, 1999] and positional (orientation and distance) reasoning about point configurations [Freksa, 1992], [Clementini et al., 1997], [Zimmermann and Freksa, 1996], [Isli and Moratz, 1999]. More recent approaches in QSR that model orientations are [Isli and Cohn, 2000], [Moratz et al., 2000]. For robot navigation, the notion of path is central [Latombe, 1991] and requires the representation of orientation and distance information [Röfer, 1999]. Since we are especially interested in qualitative calculi suitable for robot navigation we developed a positional calculus

for this task. The calculus is based on results of psycholinguistic research on reference systems. We compare the new calculus with the simpler flip-flop calculus. We can demonstrate that even if the flip-flop calculus has stronger formal properties the new calculus is better suited for certain applications in robot navigation.

2 Qualitative Representation of Relative Position

Positional calculi are influenced by results of psycholinguistic research in the field of reference systems. These findings are presented by Thora Tenbrink in the article of Moratz, Tenbrink, Fischer, and Bateman in this volume [Moratz et al., 2002]. The results point to three different options to give qualitative descriptions of spatial arrangements of objects which are labeled by Levinson [Levinson, 1996] as *intrinsic*, *relative*, and *absolute*.

We can find examples for all three options of reference systems in the QSR literature. An intrinsic reference system was used in the dipole calculus [Schlieder, 1995], [Moratz et al., 2000]. Relative reference systems in QSR were introduced by Freksa [Freksa, 1992]. Andrew Frank's cardinal direction calculus corresponds to an absolute reference system [Frank, 1991], [Ligozat, 1998].

Qualitative position calculi can be viewed as computational models for projective relations in relative reference systems. To model projective relations (like "left", "right", "front", "back") in relative reference systems, all objects are mapped onto the plane $\mathcal{D}$. The mapping of an object O onto the plane $\mathcal{D}$ is called $p_{\mathcal{D}}(O)$. The center μ of this area can be used as point-like representation O' of the object O : $O' = \mu(p_{\mathcal{D}}(O))$. Using this abstraction we will henceforth consider only point-like objects in the 2D-plane.

Figure 1 shows a simple model for the left/right-dichotomy in a relative reference system given by *origin* and *relatum* (corresponding to Levinson's terminology). Origin and relatum define the reference axis. It partitions the surrounding space in a left/right-dichotomy. The spatial relation between the reference system and the *referent* is then described by naming the part of the partition in which the referent lies. In the configuration depicted in Figure 1 the referent lies to the *left*¹ of the relatum as viewed from the origin.

This scheme ignores configurations in which the referent is positioned on the reference axis. Freksa [Freksa, 1992] used a partition that splits these configurations into three sets: the referent then is either behind the relatum, at the same position like the relatum or in front of the relatum. Ligozat [Ligozat, 1993] subdivided the arrangements with the relatum or in front of the relatum in the cases where the referent is between relatum and origin, at the same position as the origin, or behind the origin. We obtain then the partition shown on Figure 2. Ligozat calls this calculus the flip-flop calculus. For a compact notation we use abbreviations as relation symbols.

For A , B , and C as origin, relatum, and referent, Figure 3 shows point configurations and their qualitative descriptions, respectively. Isli and Moratz [Isli and Moratz, 1999] introduced two additional configurations in which origin and relatum have exactly the

¹ The natural language terms used here are meant to improve the readability of the paper. For issues of using QSR representations for modeling natural language expressions please refer to the article of Moratz, Tenbrink, Fischer and Bateman in this volume [Moratz et al., 2002].

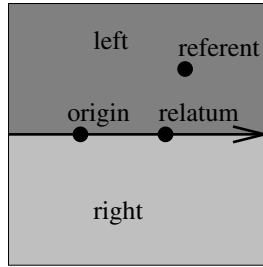


Fig. 1. The left/right-dichotomy in a relative reference system

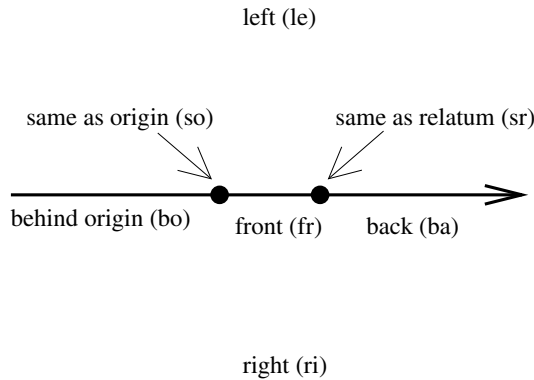


Fig. 2. Adding relations for referents on the reference axis

same location. In one of the configurations the referent has a different location, this relation is called **dou** (for double point). The configuration with all three points at the same location is called **tri** (for triple point). A system of qualitative relations which describe all the configurations of the domain and do not overlap is called jointly exhaustive and pairwise disjoint (JEPD).

The simple flip-flop calculus models “front” and “back” only as linear acceptance regions. Vorwerf et al. [Vorwerf et al., 1997] showed empirically that a cognitive adequate model for projective regions needs acceptance regions for “front” and “back” which have a similar extent as “left” and “right”. Freksa’s single cross calculus [Freksa, 1992] has this feature (see Figure 4). The front region consists of “left/front” and “right/front”, the left region consists of “left/front” and “left/back”. The intersection of both regions models the left/front relation.

The calculus we will now present is derived from the single cross calculus but makes finer distinctions. These finer distinctions are motivated by the application scenario dealing with route graphs presented at the end of our paper. The partition of the calculus is shown in Figure 5.

The letters f, b, l, r, s, d, c stand for front, back, left, right, straight, distant, close, respectively. The terms front, back, etc. are given for mnemonic purposes. The use of the TPCC relations in natural language applications is shown in this volume in an

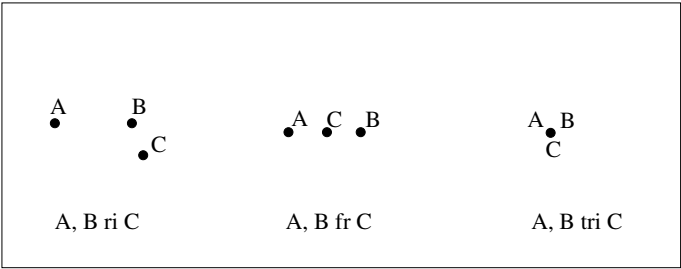


Fig. 3. Examples of point configurations and their expressions in the flip-flop calculus. We use an infix notation where the reference system consisting of origin and relatum is in front of the relation symbol and the referent is behind the relation symbol.

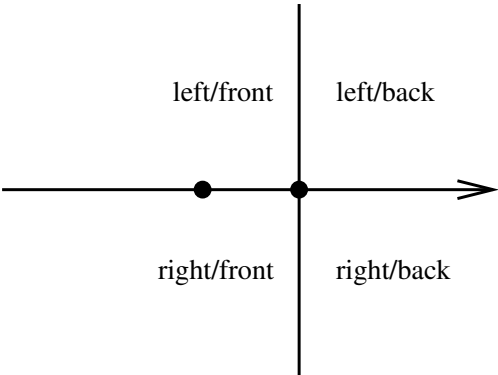


Fig. 4. The single cross calculus

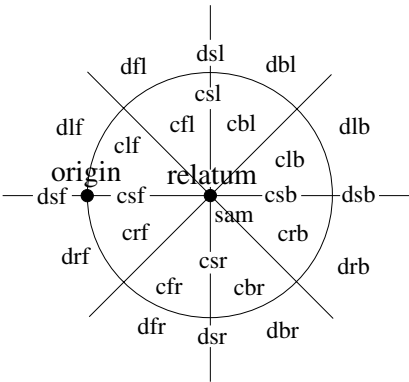


Fig. 5. The reference system used by the TPCC calculus

article by Moratz, Tenbrink, Fischer and Bateman [Moratz et al., 2002]. They use the TPCC relations for natural human robot interaction. The configuration in which the referent is at the same position as the relatum is called **sam** (for “same location”). The two special configurations in which origin and relatum have the same location **dou**, **tri** are also base relations of this calculus. This system of qualitative spatial relations and the inference rules described in the next section is called *Ternary Point Configuration Calculus* (TPCC). To give a precise, formal definition of the relations we describe the corresponding geometric configurations on the basis of a Cartesian coordinate system represented by $\mathbb{R}^2$. First we define the special cases for $A = (x_A, y_A)$, $B = (x_B, y_B)$ and $C = (x_C, y_C)$.

$$\begin{aligned} A, B \text{ dou } C &:= x_A = x_B \wedge y_A = y_B \wedge (x_C \neq x_A \vee y_C \neq y_A) \\ A, B \text{ tri } C &:= x_A = x_B = x_C \wedge y_A = y_B = y_C \end{aligned}$$

For the cases with $A \neq B$ we define a relative radius $r_{A,B,C}$ and a relative angle $\phi_{A,B,C}$:

$$\begin{aligned} r_{A,B,C} &:= \frac{\sqrt{(x_C - x_B)^2 + (y_C - y_B)^2}}{\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}} \\ \phi_{A,B,C} &:= \tan^{-1} \frac{y_C - y_B}{x_C - x_B} - \tan^{-1} \frac{y_B - y_A}{x_B - x_A} \end{aligned}$$

Then we have the following spatial relations:

$$\begin{aligned} A, B \text{ sam } C &:= r_{A,B,C} = 0 \\ A, B \text{ csb } C &:= 0 < r_{A,B,C} < 1 \wedge \phi_{A,B,C} = 0 \\ A, B \text{ dsb } C &:= 1 \leq r_{A,B,C} \wedge \phi_{A,B,C} = 0 \\ A, B \text{ clb } C &:= 0 < r_{A,B,C} < 1 \wedge 0 < \phi_{A,B,C} \leq \pi/4 \\ A, B \text{ dlb } C &:= 1 \leq r_{A,B,C} \wedge 0 < \phi_{A,B,C} \leq \pi/4 \\ A, B \text{ cbl } C &:= 0 < r_{A,B,C} < 1 \wedge \pi/4 < \phi_{A,B,C} < \pi/2 \\ A, B \text{ dbl } C &:= 1 \leq r_{A,B,C} \wedge \pi/4 < \phi_{A,B,C} < \pi/2 \\ A, B \text{ csl } C &:= 0 < r_{A,B,C} < 1 \wedge \phi_{A,B,C} = \pi/2 \\ A, B \text{ dsl } C &:= 1 \leq r_{A,B,C} \wedge \phi_{A,B,C} = \pi/2 \\ A, B \text{ cfl } C &:= 0 < r_{A,B,C} < 1 \wedge 1/2 \pi < \phi_{A,B,C} < 3/4 \pi \\ A, B \text{ dfl } C &:= 1 \leq r_{A,B,C} \wedge 1/2 \pi < \phi_{A,B,C} < 3/4 \pi \\ A, B \text{ clf } C &:= 0 < r_{A,B,C} < 1 \wedge 3/4 \pi \leq \phi_{A,B,C} < \pi \\ A, B \text{ dlf } C &:= 1 \leq r_{A,B,C} \wedge 3/4 \pi \leq \phi_{A,B,C} < \pi \\ A, B \text{ csf } C &:= 0 < r_{A,B,C} < 1 \wedge \phi_{A,B,C} = \pi \\ A, B \text{ dsf } C &:= 1 \leq r_{A,B,C} \wedge \phi_{A,B,C} = \pi \\ A, B \text{ crf } C &:= 0 < r_{A,B,C} < 1 \wedge \pi < \phi_{A,B,C} \leq 5/4 \pi \\ A, B \text{ drf } C &:= 1 \leq r_{A,B,C} \wedge \pi < \phi_{A,B,C} \leq 5/4 \pi \end{aligned}$$

$$\begin{aligned}
A, B \text{ cfr } C &:= 0 < r_{A,B,C} < 1 \wedge 5/4 \pi < \phi_{A,B,C} < 3/2 \pi \\
A, B \text{ dfr } C &:= 1 \leq r_{A,B,C} \wedge 5/4 \pi < \phi_{A,B,C} < 3/2 \pi \\
A, B \text{ csr } C &:= 0 < r_{A,B,C} < 1 \wedge \phi_{A,B,C} = 3/2 \pi \\
A, B \text{ dsr } C &:= 1 \leq r_{A,B,C} \wedge \phi_{A,B,C} = 3/2 \pi \\
A, B \text{ cbr } C &:= 0 < r_{A,B,C} < 1 \wedge 3/2 \pi < \phi_{A,B,C} < 7/4 \pi \\
A, B \text{ dbr } C &:= 1 \leq r_{A,B,C} \wedge 3/2 \pi < \phi_{A,B,C} < 7/4 \pi \\
A, B \text{ crb } C &:= 0 < r_{A,B,C} < 1 \wedge 7/4 \pi \leq \phi_{A,B,C} < 2 \pi \\
A, B \text{ drb } C &:= 1 \leq r_{A,B,C} \wedge 7/4 \pi \leq \phi_{A,B,C} < 2 \pi
\end{aligned}$$

There are cases in which we only have coarser spatial knowledge or in which we are at the border of a segment of the partition and cannot decide safely due to measurement errors. Then we use sets of the above defined relations to denote disjunctions of relations. Figure 6 shows a situation where it is not sensible to decide visually between the alternatives $A, B \text{ clb } C$ and $A, B \text{ cbl } C$. Such a configuration is described by the relation $A, B \text{ (cbl, clb) } C$.

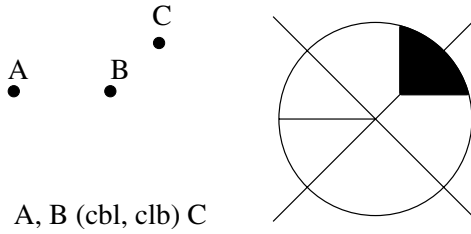


Fig. 6. Coarser spatial knowledge

3 Deductive Reasoning about Relative Positional Information

In the last section we defined relations between triples of points on the 2D-plane. Now we define a set of unary and binary operations that allow to deduce new relations about point sets from given relations about these points. Unary operations (transformations) use relations about three points to deduce a relation which holds for a permuted sequence of the same points. Binary operations (compositions) deduce information from two relations which have two points in common (the set consists of four points). The result then is a relation about one of the common points and the two other points.

3.1 Permutations

Because we have three arguments, we have $3! = 6$ possible ways of arranging the arguments for a transformation. Following Zimmermann and Freksa

[Zimmermann and Freksa, 1996] we use the following terminology and symbols to refer to these permutations of the arguments (a,b : c):

term	symbol	arguments
identical	Id	a,b : c
inversion	Inv	b,a : c
short cut	Sc	a,c : b
inverse short cut	SCI	c,a : b
homing	HM	b,c : a
inverse homing	HMI	c,b : a

The transformation tables for the flip-flop calculus are presented in Isli and Moratz [Isli and Moratz, 1999]. We therefore present here only the transformation table for the TPCC calculus on table 8. In contrast to the flip-flop calculus the TPCC calculus is not closed under the transformations. That means that results of a transformation can constitute proper subsets of the base relations. Since we need many sets of relations as results of transformed relations we introduce here an iconic notation of the relations which makes the presentation more compact:

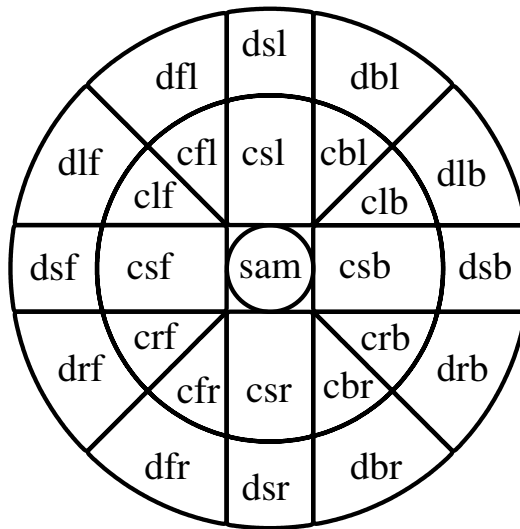


Fig. 7. Iconic Representation for TPCC-Relations

The segments corresponding to a relation are presented as filled segments. Unions of relation then simply have several segments filled. The reference axis and the dividing lines between left, right, front and back are also presented in the icon to make the visual identification of the relation symbol easier. The iconic representation is easier to translate into its semantic content (the denoted spatial point configuration) compared with a representation that uses the textual relation symbol. And unions can be expressed in a compact way.

ID															
INV															
SC															
SCI															
HM	dou														
HMI	dou														

Fig. 8. Permutation Table for TPCC-Relations

In order to reduce the size of the table trivial cases for **dou** and **tri** are omitted. Symmetric cases can be derived using a reflection operation (reflection on an axis). The results of $SC(\mathbf{dsf})$ and $SCI(\mathbf{dsf})$ also include **dou** as a result.

3.2 Composition

With ternary relations, one can think of different ways of composing them. However there are only a few ways to compose them in a way such that we can use it for enforcing local consistency [Scivos and Nebel, 2001]. In trying to generalize the path-consistency algorithm [Montanari, 1974], we want to enforce 4-consistency [Isli and Cohn, 2000]. We use the following (strong) composition operation:

$$\forall A, B, D : A, B (r_1 \diamond r_2) D \leftrightarrow \exists C : A, B (r_1) C \wedge B, C (r_2) D$$

The composition table for the flip-flop calculus is presented in Isli and Moratz [Isli and Moratz, 1999].

Unfortunately, the TPCC calculus is not closed under strong composition. For that reason we can not directly enforce 4-consistency. But we can define a weak composition operation $r_1 \diamond r_2$ of two relations r_1 and r_2 . It is the most specific relation such that:

$$\forall A, B, D : A, B (r_1 \diamond r_2) D \leftarrow \exists C : A, B (r_1) C \wedge B, C (r_2) D$$

While using the weak composition we can not enforce 4-consistency we still get usefull inferences. We use this weak composition for inferences in the application scenario in section 4.

The table for weak composition of TPCC relations is shown in figure 9. The first operand determines the row, the second operand the column. Again the table omits entries which can be found by reflection in order to reduce the size of the table. And the trivial cases for **dou** and **tri** are omitted.

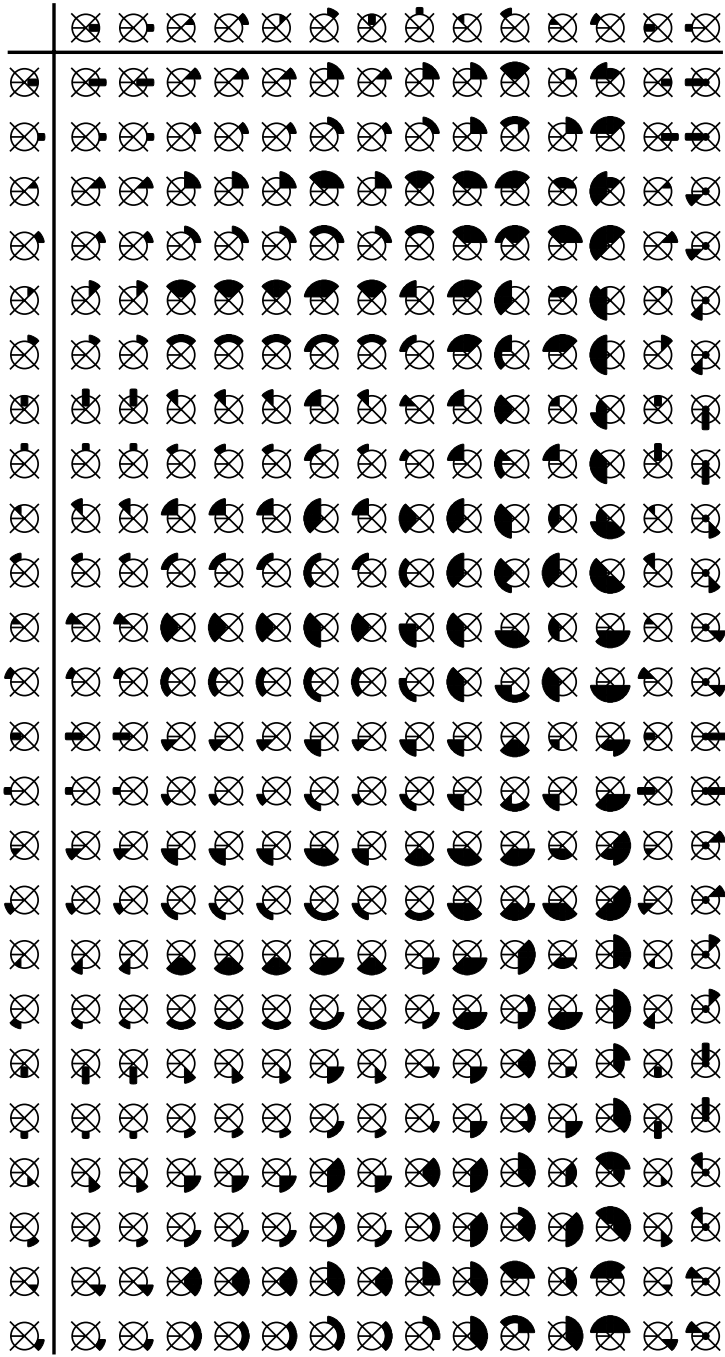


Fig. 9. Composition of TPCC-Relations

3.3 Constraint-Based Reasoning

The standard method for reasoning with relation algebras is to use Ladkin and Reinefeld's algorithm [Ladkin and Reinefeld, 1992] that uses backtracking employing the path-consistency algorithm as forward checking method. This scheme was extended by Isli and Cohn [Isli and Cohn, 2000] for ternary relation algebras. It can then easily be applied to the flip-flop calculus.

A prerequisite to using the standard constraint algorithms is to express the calculi in terms of relation algebras in the sense of Tarski [Ladkin and Maddux, 1994]. But since the TPCC-Calculus is not closed under the transformations and under the composition we can not use this scheme. However, simple path-based inferences can be performed using the following scheme. The two last relations of a path are composed. Then the reference system is incrementally moved towards the beginning of the path in form of a backward chaining.

For the detection of cyclic paths a reference system consisting of a path segment in the middle of the path is appropriate. Then the relative position of the points in both directions is derived and compared using an inversion operation (see appendix A for an example).

It can be proven that reasoning with the TPCC relations is in **PSPACE**. The idea of the proof sketch is as follows. The algebraic semantics of the relations implies that reasoning problems in the TPCC calculus can be expressed as equalities over polynomials with integer coefficients. Systems of such equalities can be solved using polynomial space [Renegar, 1992].

4 Path-Based Reasoning in Route Graphs

The flip-flop calculus and the TPCC calculus can be used to integrate local and survey knowledge about the spatial environment of an agent. Local knowledge can be sensorically acquired from one fixed point in space. Survey knowledge is an abstraction that integrates a number of local perceptions into a coherent whole. The local perceptions are typically acquired in a sequence during an exploration process. The accumulated local assessments of qualitative configurations have local frames of reference. The integration process needs to reason about the position of the salient objects in a global reference. Path integration can serve as a means to achieve the accumulation of local orientation information. The problem of detecting cyclic paths in a route graph is a sample application which we present to compare the coarse and the finer calculus in a typical application scenario.

From a qualitative viewpoint a path can be viewed as a sequence of qualitative positions. The positions are discriminated with respect to the environment. Therefore both calculi can be used to describe and to reason about paths. Qualitative positional reasoning about paths is used for robot navigation in the approach of Sogo and in the approach of Musto [Sogo et al., 1999] [Musto et al., 1999]. In our sample application the environment consists of a route graph [Werner et al., 1998]. The task is to derive a global map from locally perceived information. Reasoning from perceived local spatial arrangements about the underlying global layout of an environment is a form of abductive reasoning [Remolina and Kuipers, 2001] [Shanahan, 1996].

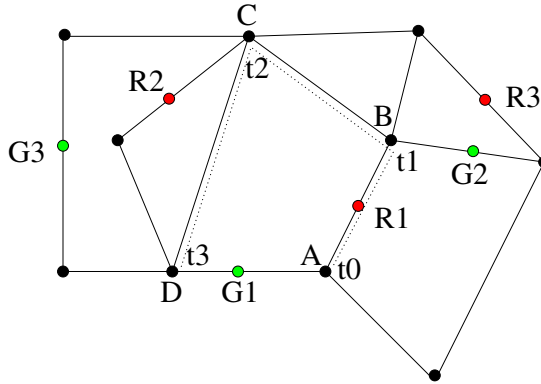


Fig. 10. A route graph

We focus here on a deductive subproblem. The problem is to decide whether two landmarks reached during an exploration can be identical due to a cycle in the path. The observations are collected at timepoints t_0 , t_1 , t_2 and t_3 (see figure 10). The local observations are expressed in both calculi. Then we test which landmarks G_1 perceived from A, G_2 , G_1 perceived from t_3 and R_1 , R_2 can be deduced to be distinct. The observations and the deductive inferences are listed in appendix A.

The result is that both calculi can deduce that G_1 and G_2 are distinct. Both calculi are correct and therefore do not deduce that G_1 perceived from A and G_1 perceived from t_3 are distinct. But only the TPCC calculus can deduce that R_1 and R_2 are distinct. Using the same reasoning scheme the TPCC calculus can also deduce that R_1 and R_3 are distinct which needs not only orientation but also distance-based reasoning. This example shows that differentiations which are useful in realistic application scenarios are supported by the new TPCC calculus. These finer (but still coarse) distinctions can not be dealt with in the mathematically more elegant flip-flop calculus.

There are applications in which even finer qualitative acceptance areas are helpful. The techniques described here can still be used. But there is obviously no way to design icons that can express these finer distinctions. And the computation of the composition table can become difficult. Then an approximation of the composition results can be used. The possibility to use even finer qualitative distinctions can be viewed as a stepwise transition to quantitative knowledge which is the topic of the next subsection.

4.1 Comparison with a Quantitative Approach for Interval-Based Spatial Reasoning

The simplest and most common strategy to deal with coarse knowledge is to treat it as if it were precise metrical knowledge. Then the user has to rely on his good luck that all derived conclusions are valid. Compared to that unsafe approach qualitative spatial reasoning is *safe* because it only derives correct information as long as the input information is correct. This technical argument for QSR leads to the question whether QSR is the only way to do safe spatial reasoning. In scalar or one-dimensional domains

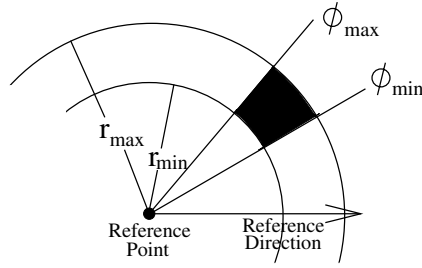


Fig. 11. A distance/orientation-interval and its parameters

interval-based reasoning serves the purpose of a safe quantitative alternative to qualitative reasoning. Therefore we now present a straightforward quantitative approach which is based on distance/orientation-intervals.

A distance/orientation-interval (DOI) uses a point and a reference direction as anchor and has four additional parameters r_{min} , r_{max} , ϕ_{min} and ϕ_{max} (see figure 11).

These quantitative intervals can be propagated along paths analogous to the qualitative counterparts. The respective reference directions then are determined by adjacent points on the path. The technical details of the interval propagation can be found in [Moratz, *in preparation*]. The quantitative calculus can solve all presented problems of the last section about the route graph when the observed intervals are sufficiently small.

Now we look at an example where we need a more expressive calculus. In the route graph depicted on figure 12 an agent travels from D to E via A , B , C . We model the perception of the agent like in the previous example. The agent can only perceive locations to which a direct straight link exits. Then it can not use the propagation of measured intervals to distinguish between D and E if the distance between D and E is sufficiently small.

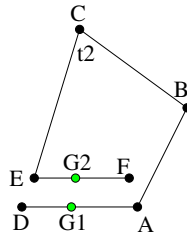


Fig. 12. Reasoning about the absence of features

We need to represent the information that seen from D to A there is no road junction at a direction differing from A . Because QSR can be seen as reasoning about space within first order logic [Isli and Cohn, 2000] [Renz and Nebel, 1999] we have *negation*, *disjunction* and *conjunction* already built in. So we can use the TPCC-Calculus to express our knowledge about the absence of a feature:

$$\forall x \in J, g \in G \neg ((g, D \text{ --- } x) \wedge \text{cn}(E, g) \wedge \text{cn}(E, x))$$

The symbol **cn** stands for the predicate *connected* (via a direct straight link). J is the set of all road junctions, G is the set of all green landmarks. Adding this logical constraint to the observations we can distinguish the road junctions D and E . We can not express this in the quantitative calculus because we have no logical operations. To extend a quantitative calculus in that direction is not a trivial task and would make it much more complex.

5 Conclusion and Perspective

We presented the new TPCC calculus for representing and reasoning about qualitative relative position information. We identified a system of 27 atomic relations between points and computed the composition table based on their algebraic semantics, which allows to apply constraint-based reasoning methods. It was demonstrated that reasoning with the TPCC relations is in PSPACE. Potential applications of the calculus are demonstrated with a small navigation example in route graphs.

In a comparison with a coarser calculus known in the literature we noticed that helpful mathematical properties are unfortunately not satisfied by the TPCC calculus. It is a matter of further studies how the framework of constraint satisfaction especially with respect to path consistency can be transferred to the new calculus.

Acknowledgement

The authors would like to thank Amar Isli, Jochen Renz, Alexander Scivos and Thora Tenbrink for interesting and helpful discussions related to the topic of the paper. And we would like to thank Sven Kröger for computing the composition table. This work was supported by the DFG priority program on Spatial Cognition.

References

- Clementini et al., 1997. Clementini, E., Di Felice, P., and Hernandez, D. (1997). Qualitative representation of positional information. *Artificial Intelligence*, 95:317–356.
- Cohn, 1997. Cohn, A. (1997). Qualitative spatial representation and reasoning techniques. In Brewka, G., Habel, C., and Nebel, B., editors, *KI-97: Advances in Artificial Intelligence*, Lecture Notes in Artificial Intelligence, pages 1–30. Springer-Verlag, Berlin.
- Frank, 1991. Frank, A. (1991). Qualitative spatial reasoning with cardinal directions. In *Proceedings of 7th Österreichische Artificial-Intelligence-Tagung*, pages 157–167, Berlin. Springer.
- Freksa, 1992. Freksa, C. (1992). Using Orientation Information for Qualitative Spatial Reasoning. In Frank, A. U., Campari, I., and Formentini, U., editors, *Theories and Methods of Spatial-Temporal Reasoning in Geographic Space*, pages 162–178. Springer, Berlin.
- Isli and Cohn, 2000. Isli, A. and Cohn, A. (2000). Qualitative spatial reasoning: A new approach to cyclic ordering of 2d orientation. *Artificial Intelligence*, 122:137–187.
- Isli and Moratz, 1999. Isli, A. and Moratz, R. (1999). *Qualitative Spatial Representation and Reasoning: Algebraic Models for Relative Position*. Universität Hamburg, FB Informatik, Technical Report FBI-HH-M-284/99, Hamburg.

- Ladkin and Maddux, 1994. Ladkin, P. and Maddux, R. (1994). On binary constraint problems. *Journal of the Association for Computing Machinery*, 41(3):435–469.
- Ladkin and Reinefeld, 1992. Ladkin, P. and Reinefeld, A. (1992). Effective solution of qualitative constraint problems. *Artificial Intelligence*, 57:105–124.
- Latombe, 1991. Latombe, J.-C. (1991). *Robot Motion Planning*. Kluwer.
- Levinson, 1996. Levinson, S. C. (1996). Frames of Reference and Molyneux's Question: Crosslinguistic Evidence. In Bloom, P., Peterson, M., Nadel, L., and Garrett, M., editors, *Language and Space*, pages 109–169. MIT Press, Cambridge, MA.
- Ligozat, 1993. Ligozat, G. (1993). Qualitative triangulation for spatial reasoning. In *COSIT 1993*, Berlin. Springer.
- Ligozat, 1998. Ligozat, G. (1998). Reasoning about cardinal directions. *Journal of Visual Languages and Computing*, 9:23–44.
- Montanari, 1974. Montanari, U. (1974). Networks of constraints: Fundamental properties and applications to picture processing. *Information Sciences*, 7:95–132.
- Moratz. Moratz, R. Propagation of distance-orientation intervals: Finding cycles in route graphs. *in preparation*.
- Moratz et al., 2000. Moratz, R., Renz, J., and Wolter, D. (2000). Qualitative spatial reasoning about line segments. In W., H., editor, *ECAI 2000. Proceedings of the 14th European Conference on Artificial Intelligence*, Amsterdam. IOS Press.
- Moratz et al., 2002. Moratz, R., Tenbrink, T., Fischer, F., and Bateman, J. (2002). Spatial knowledge representation for human-robot interaction. *this volume*.
- Musto et al., 1999. Musto, A., Stein, K., Eisenkolb, A., and Röfer, T. (1999). Qualitative and quantitative representations of locomotion and their application in robot navigation. In *Proceedings IJCAI-99*, pages 1067–1072.
- Randell et al., 1992. Randell, D., Cui, Z., and Cohn, A. (1992). A spatial logic based on regions and connection. In *Proceedings KR-92*, pages 165–176, San Mateo. Morgan Kaufmann.
- Remolina and Kuipers, 2001. Remolina, E. and Kuipers, B. (2001). A logical account of causal and topological maps. In *Proceedings IJCAI-2001*.
- Renegar, 1992. Renegar, J. (1992). On the computational complexity and geometry of the first order theory of the reals. part I-iii. *Journal of Symbolic Computation*, 13:3:255–352.
- Renz and Nebel, 1999. Renz, J. and Nebel, B. (1999). On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the region connection calculus. *Artificial Intelligence*, 108(1-2): 69–123.
- Röfer, 1999. Röfer, T. (1999). Route Navigation Using Motion Analysis. In Freksa, C. and Mark, D., editors, *COSIT 1999*, pages 21–36, Berlin. Springer.
- Schlieder, 1995. Schlieder, C. (1995). Reasoning about ordering. In A Frank, W. K., editor, *Spatial Information Theory: a theoretical basis for GIS*, number 988 in Lecture Notes in Computer Science, pages 341–349, Berlin. Springer Verlag.
- Scivos and Nebel, 2001. Scivos, A. and Nebel, B. (2001). Double-crossing: Decidability and computational complexity of a qualitative calculus for navigation. In *COSIT 2001*, Berlin. Springer.
- Shanahan, 1996. Shanahan, M. (1996). Noise and the common sense informatic situation for a mobile robot. In *Proceedings AAAI-96*.
- Sogo et al., 1999. Sogo, T., Ishiguro, H., and Ishida, T. (1999). Acquisition of qualitative spatial representation by visual observation. In *Proceedings IJCAI-99*, pages 1054–1060.
- Vorwerg et al., 1997. Vorwerg, C., Socher, G., Fuhr, T., Sagerer, G., and Rickheit, G. (1997). Projective relations for 3d space: Computational model, application, and psychological evaluation. In *AAAI'97*, pages 159–164.
- Werner et al., 1998. Werner, S., Krieg-Brückner, B., and Herrmann, T. (1998). Modelling Navigational Knowledge by Route Graphs. In Freksa, C., Habel, C., and Wender, K. F., editors,

Spatial Cognition II, Lecture Notes in Artificial Intelligence, pages 295–317. Springer-Verlag, Berlin.

Zimmermann and Freksa, 1996. Zimmermann, K. and Freksa, C. (1996). Qualitative spatial reasoning using orientation, distance, path knowledge. *Applied Intelligence*, 6:49–58.

Appendix A: Inferences for the Route Graph Example

First we use the flip-flop calculus for representation and reasoning in the route graph example. We have the following observations at timepoints t :

$$\begin{aligned}
 t_0 : \quad & e_2(B), e_1(A) \quad \text{ri} \quad g_1(G_1) & (1) \\
 t_1 : \quad & e_3(C), e_2(B) \quad \text{ri} \quad e_1(A) & (2) \\
 t_1 : \quad & e_1(A), e_2(B) \quad \text{ri} \quad g_2(G_2) & (3) \\
 t_2 : \quad & e_2(B), e_3(C) \quad \text{le} \quad e_4(D) & (4) \\
 t_3 : \quad & e_3(C), e_4(D) \quad \text{le} \quad g_3(G_1) & (5)
 \end{aligned}$$

The observed crossing points are denoted e_1, e_2, e_3, e_4 . The corresponding points on figure 10 are appended in brackets. Please note that landmark G_1 gets a new internal label g_3 by the exploring agent when observed the second time. Using these observations we make the following inferences on a syntactical basis using the operations defined in section 3:

We apply the inversion tranform to equation (1):

$$e_1(A), e_2(B) \quad \text{le} \quad g_1(G_1) \quad (6)$$

Now we test whether g_1 and g_2 can be the same landmark. Therefore we make the assumption that g_1 and g_2 are the same point. The intersection operation between qualitative spatial relations about the same points is simply the set theoretic intersection about the sets of atomic relations associated with each of the two relations. Since we made the assumption that g_1 and g_2 are the same we can apply the intersection operation an equations (3) and (6). The intersection is empty. The empty set as qualitative spatial relation corresponds semantically to an impossible spatial arrangement of points. Then we can deduce a contradiction from our assumption that g_1 and g_2 are the same point. It follows that g_2 is different from g_1 .

For comparism we use the TPCC calculus for the same example. The observations and the inferences are:

$$t0: \quad e2(B), e1(A) \quad \text{⊗} \quad g1(G1) \quad (1)$$

$$t1: \quad e3(C), e2(B) \quad \text{⊗} \quad e1(A) \quad (2)$$

$$t1: \quad e1(A), e2(B) \quad \text{⊗} \quad g2(G2) \quad (3)$$

$$t2: \quad e2(B), e3(C) \quad \text{⊗} \quad e4(D) \quad (4)$$

$$t3: \quad e3(C), e4(D) \quad \text{⊗} \quad g3(G1) \quad (5)$$

$$(1) \xrightarrow{\text{inversion}} e1(A), e2(B) \quad \text{⊗} \quad g1(G1) \quad (6)$$

$$(3), (6) \xrightarrow{\text{empty intersection}} g2 \text{ different from } g1$$

$$(2), (1) \xrightarrow{\text{composition}} e3(C), e2(B) \quad \text{⊗} \quad g1(G1) \quad (7)$$

$$(4), (5) \xrightarrow{\text{composition}} e2(B), e3(C) \quad \text{⊗} \quad g3(G1) \quad (8)$$

$$(7) \xrightarrow{\text{inversion}} e2(B), e3(C) \quad \text{⊗} \quad g1(G1) \quad (9)$$

$$(8), (9) \xrightarrow{\text{non-empty intersection}} g3 \text{ potentially identical with } g1$$

Interpretation of Intentional Behavior in Spatial Partonomies

Christoph Schlieder and Anke Werner

Technologie-Zentrum Informatik, Universität Bremen, Postfach 330440

28334 Bremen, Germany
{cs, anke}@tzi.de
<http://www.tzi.de/>

Abstract. Information services that are accessed by mobile computers need to compensate the limited potential for direct interaction by some mechanism which analyzes the intentions of the user. Location-aware information services, for instance, take the decision about what is relevant to the user on ground of information about the user's spatial location. We show that if the regions of the geographic space in which the user moves are structured hierarchically by partonomies a context problem arises. To resolve the problem, not only the user's location but also his motion must be taken into account. We propose a location model that supports inferring intentional behavior in spatial partonomies from motion patterns and describe the architecture of the corresponding modeling framework.

1 Intentional Behavior in Geographic Space

Location-aware services pioneered by researchers at the Xerox Parc Laboratory under the vision of ubiquitous computing (see e.g. Schilit & al., 1993) exploit the idea that the intentions of a human agent can be inferred from information about his current location. This is a valid assumption in certain cases. However, intentional behavior often correlates with complex motion patterns rather than just with location. A further challenge for location modeling comes into play when the user's mental representation of space is considered. From psychological research it is known that region-based representations of geographic space tend to be organized hierarchically by part-of relations (Hirtle, 1995). Thus, intentional spatial behavior seems intrinsically bound to what AI research has called spatial partonomies (e.g. Winston & al., 1987; Davis, 1990). For ubiquitous computing, this raises the problem of finding a suitable location model for identifying the intentions of a user who is moving in an environment structured by partonomies.

This paper reports on a particular instance of this problem which we encountered in the Tourserv project¹. Within the scope of this project, a service platform has been built that provides regional information and navigation support to tourists. A pilot

¹ Funded by the European Union IST-1999-20414

system is currently implemented for the Italian ski resort of Scopello near Milan. It will permit tourists to use mobile devices, such as PDA or smart phones, to get optimal support during their skiing, mountaineering, or hiking activities. The tourist's actual position is gained by GPS and this information is used to guide proactive information presentation and navigation support.

The class of spatial positions which must be distinguished, the type of information services offered, and especially the relationship between both vary from one problem domain to another. In a ski resort, the queuing area in front of a ski lift constitutes a relevant spatial region. If a person is localized in this region, chances are high that this person is intending to use the ski lift. Based on this hypothesis, the system can decide which information service to offer (e.g. temperature and wind speed at the top end of the ski lift). The scope of such interpretation rules is limited: in general, they cannot be transferred to another domain, e.g. a location-aware information service in a museum.

Thus, the main lesson learned from the Tourserv project is the need for a modeling framework which permits the system developer to describe interpretation rules that map the observed spatial behavior of the user onto supporting information services. This paper describes a modeling framework which provides the means to design services that can be adapted to other application areas with much less effort than the domain-specific solution implemented in the Tourserv system. As another lesson from the project, it became clear that spatial position by itself is a poor predictor for the user's intentions. This issue cannot be resolved by increasing the precision of measurements. Positioning technologies like GPS are able to provide sufficiently exact information for navigation purposes but they do not solve the problem of identifying which spatial context is relevant for the user.

The rest of the paper presents a modeling framework and an interpretation mechanism for spatial behavior in environments structured by paronomies. It is organized as follows. Section 2 introduces the layered approach to the interpretation of spatial behavior. Two basic problems that any interpretation mechanism must address are discussed in sections 3 and 4: the spatial context problem and the motion segmentation problem. It is shown how the paronomic structure of the spatial environment can be used to solve both problems. In section 5, the modeling framework is presented which consists of a representational formalism for spatial behavior encoding partitioned motion patterns and an interpretation mechanism for these patterns. We conclude with a discussion of related work and an outlook on future research in section 6.

2 From Behavior to Services: A Layered Approach

Location-aware services are based on the idea of interpreting the simplest possible type of spatial behavior: being located at a certain place. Places are generally described with reference to a *semantic location model* rather than by a position in a geographic reference system (e.g. UTM coordinates used by GPS). Such a semantic location model specifies a number of spatio-thematic regions, i.e. spatial regions that possess thematic relevance in the application domain.

The behavior of a person using the information system is characterized by the spatio-thematic region the person is located in during query time. Each of these regions is associated with a specific information service. This association can be hard-coded or implemented by a rule-based approach in which interpretation rules specify the behavior-service mapping. A simple interpretation rule from the example domain is *behavior(ski-lift-queuing-area) → service(weather-broadcast)*. The rule-based design has the advantage to permit building generic information systems that can be adapted to different application domains simply by exchanging the set of rules.

User Intentions as Additional Modeling Layer

Although the mapping of behavior to services via rules increases portability, more modeling flexibility is needed in domains where the spatio-thematic regions or the information services tend to change. The main reason for change in the spatio-thematic regions is the introduction of new facilities (e.g. restaurants) and of new types of position sensors (e.g. points of sale reporting payments). Typically, the information services offered to the user change at an even higher rate since almost every new source of information or new method of presentation generates an improved service.

In contrast, the user's needs, insofar as they translate into intentions, remain rather stable. A person forms the intention to use a ski lift independently of the type of position sensor that monitors his spatial behavior or the information services that are currently available. To increase flexibility and knowledge reuse, our architecture introduces intentions as an additional descriptive level beside the behavioral and the service level. As a consequence, two different kinds of mapping rules are needed. First, rules interpreting behavior in terms of intentions, second, rules associating intentions with services. The simple interpretation rule given above expands into two rules as shown in Fig.1.

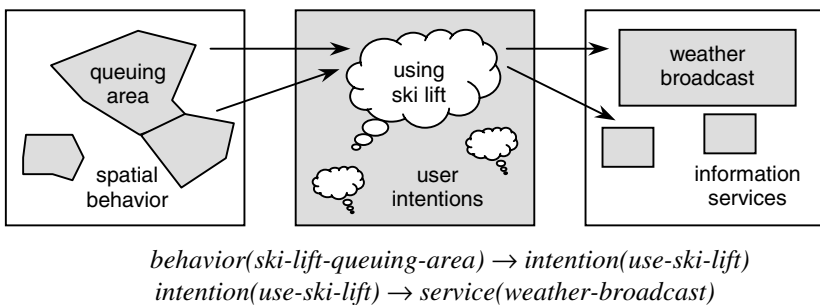


Fig. 1. Intentions as an intermediate level in modeling behavior-service mappings

Only a limited number of user intentions is relevant to an application domain and needs to be distinguished by the system. Typically, different types of spatial behavior are mapped onto the same intention whereas the same intention may be associated with different information services. For this reason, the two-rule-set approach uses generally less rules than the single-rule-set approach. If n_i is the number of behavior

rules for an intention i , and m_i the number of service rules for that intention, then a single-rule-set approach has to specify $n_i \cdot m_i$ rules where the two-rule-set approach needs only $n_i + m_i$ rules. More important, the two-rule-set design supports knowledge reuse because it permits to confine changes in most cases to either the behavior or the service rule set. If, for instance, a new type of position sensor is introduced, a new spatio-thematic region is created which only affects the behavior rule set.

3 The Spatial Context Problem

A position on a digital map typically corresponds not to a single region but to a hierarchy of regions. The tourist located at the queuing area of a ski lift in the resort of Scopello is also located in the commune of Scopello, in the valley of Varese, and in Italy. Depending on the tourist's intentions, any of these regions can become the focus of relevance for information services. Partonomies are the result of recursively applying the spatial part-of relation to describe the decomposition of wholes into parts, i.e. regions into subregions. In our approach we make use of the representation for spatial partonomies described by Schlieder & al. (2001) in a geographic information system (GIS) context.

Encoding Spatial Partonomies

Different types of spatial part-of relations can be distinguished. To define these types, we assume that the regions are encoded as polygons and that each polygon is a closed set of points, i.e. edges and vertices belong to the polygon. The restriction to polygons is motivated by the fact that spatio-thematic regions are usually managed by a GIS using a vector representation for geographic objects.

If we consider polygons $P_1, \dots, P_n$ that are contained in a part of the plane bounded by a polygon P , three types of arrangements of the polygons within the containing polygon P can be distinguished.

- (1) *polygonal covering* where $P_1 \cup \dots \cup P_n = P$. The polygons cover the containing polygon. In general, they will overlap.
- (2) *polygonal patchwork* where for all $i \neq j$ from $\{1, \dots, n\}$ interior $(P_i \cap P_j) = \emptyset$. The polygons are either disjoint or intersect only in edges and/or vertices.
- (3) *polygonal tessellation* which is a polygonal covering that also forms a polygonal patchwork.

A set of polygons $\{P_1, \dots, P_n\}$ that constitutes a covering, patchwork, or tessellation of P forms a *decomposition* of P . We introduce the following notation for the spatial part-of relation. If P_i is one of the polygons of a decomposition D of P we say that P_i is a *spatial part of P with respect to D* and write $P_i \leq_D P$. The subscript denoting the decomposition can be omitted whenever it becomes clear from the context which decomposition is meant. Note however, that the spatial part-of relation is not a binary but a ternary relation with a polygonal part, a decomposition, and a polygonal

whole as arguments. Recursive decompositions arise from decomposing the parts of a decomposition. The spatial part-of relation of a recursive decomposition R is defined as the transitive closure of the union of the spatial part-of relations of all its constituting decompositions: $P_0 \leq_R P_n$ if there are decompositions $D_1, \dots, D_n$ from R with $P_0 \leq_{D_1} P_1, \dots, P_{n-1} \leq_{D_n} P_n$.

The *decomposition tree* describes the recursive decomposition of a polygon into polygonal parts by a tree whose edges encode the spatial part-of relation at each level of decomposition. The nodes of the decomposition tree denote polygons and are labelled with one of the following decomposition types: patchwork, covering, tessellation, non-decomposed. For the purpose of our analysis, decomposition trees provide an adequate formalization of *partonomies*. Therefore, we will use both terms synonymously. The partonomies considered in the following are all derived from patchwork or tessellation decompositions.

For polygonal tessellations, it is possible to formalize both metric (denoting distance), ordinal (denoting directions), and topological (denoting neighborhoods) spatial relations (Schlieder, 1996). These can be represented using graph-theoretical constructs like neighborhood- and connection graphs (Schlieder et al., 2001). Given the existence of valid quantitative GIS or CAD data, the automatic or semi-automatic creation of such qualitative models is straight-forward.

Determining the Relevant Spatial Context

We analyze the problem of spatial context and present our solution for that problem using a simplified application scenario. In this scenario, a tourist explores an art museum. He is assisted by a mobile device connected to the museum's tourist information system. The spatial context problem arises whenever the tourist's location is not part of a single spatio-thematic region but part of multiple, hierarchically stacked spatio-thematic regions. This situation occurs frequently as the museum is structured by a partonomy consisting of a number of wings each of which is subdivided into several exhibition rooms which in turn hold a number of exhibits. Part of the corresponding decomposition tree can be seen on the right side in Fig. 2.

Localization in the semantic location model represented by the decomposition tree is ambiguous in the sense that the position of a visitor corresponds generally to several spatio-thematic regions. In that case, more than one behavior rule and more than one service rule are applicable. This problem appears, of course, also with single-rule-set architectures. The standard solution for the problem consists in adopting a *smallest-containing-region strategy*: Associate the visitor's position v with the spatio-thematic region P from the decomposition tree D if v lies within P and v does not lie within any other region P' with $P' \leq_D P$.

This strategy has been successfully applied in mobile working environments with shallow partonomies where a well-defined workflow induces a strict association of leaf-node regions in the decomposition tree with specific working tasks (Kirste, Rieck, and Schumann, 1997). In environments with a more complex partonomic structure, such as a museum or a ski resort, the strategy is likely to fail. Consider the case where the visitor of a museum enters the spatio-thematic region associated with a particular exhibit while quickly walking through the room to gain the next room.

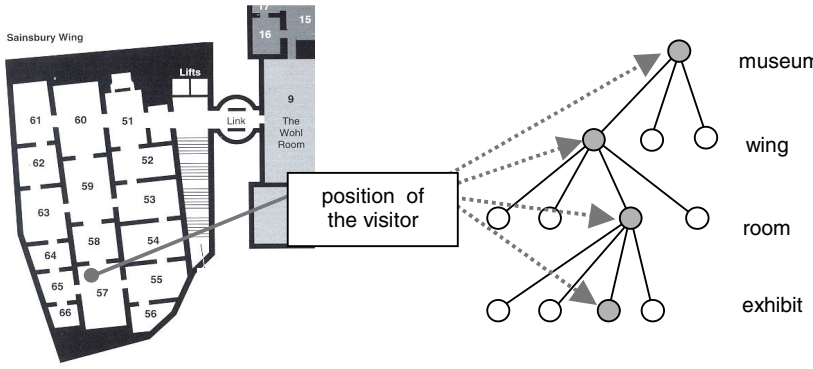


Fig. 2. Ambiguous location in the museum parthood

According to the smallest-containing-region strategy, the region belonging to the exhibit will be identified as the relevant spatial context. Crossing the room, the visitor enters several regions that belong to exhibits all of which are identified as relevant contexts. The system will prompt the visitor with (unwanted) information about these exhibits instead of displaying, for instance, a map of the wing the room is located in.

In principle, the solution to the problem is straight-forward. Once that the visitor's intention has been interpreted correctly as that of crossing the room, the relevant spatial context is easily identified as the room (and not the exhibit). Obviously, the visitor's position provides too little information for interpreting spatial behavior in this case. It is rather the motion pattern of the visitor that must be considered. In the example illustrated in Fig. 3, the visitor's intentions are interpreted and expressed in terms of the relevant spatio-thematic region at 7 positions along his path. To keep notation simple, we assume that only a unique intention exists for each of the regions: a visitor having intention *A* intends to get informed about room *A*, similarly, intention 1 is the intention to get informed about exhibit 1.

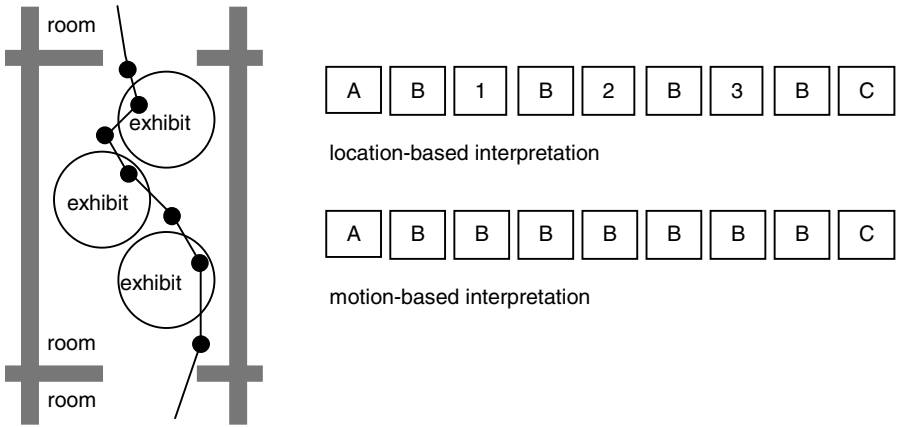


Fig. 3. Location-based and motion-based interpretation of spatial behavior

The smallest-containing-region strategy has the advantage of solving the spatial context problem with behavior rules that refer to spatial location only. Unfortunately, this simple location-based interpretation leads to unintuitive results. In order to correctly identify the user's intention in the example, a motion-based approach is needed that takes into account the duration of the visitor's stay in a region and, in more complex cases, also the shape of the path. Such behavior rules are more complex in the sense that they require a representational formalism for expressing motion patterns. We present an adequate formalism in section 5. Note that in a two-rule-set architecture, we can pass from a location-based approach to a motion-based approach by just adapting the behavior rules since only the interpretation of the user's spatial behavior is concerned.

In some application domains, even motion-based approaches will fail simply because spatial behavior on its own does not provide sufficient information to predict the user's intention. Just consider an extreme case: from observing someone walking to a library's information desk we cannot infer what book he is going to ask for. However, location-aware services aim at use cases with much less complex intentions. They work perfectly well, for instance, when it comes to decide whether the number to be dialed for calling a taxi is that of the user's home town or that of another place. As we have argued, the motion-based approach covers an even vaster range of applications

4 The Motion Segmentation Problem

From the perspective of the information service, the user's spatial behavior reduces to a time series of measurements of his spatial position, the *motion pattern*. Motion patterns constitute the raw data that our motion-based approach starts with. The central problem consists in interpreting these patterns with respect to the user's intentions, that is, in translating a motion pattern into a time series of intentions which we call *intention sequence*.

Behavioral Specificity Assumption and Partonomic Segmentation

The interpretation process is simplified when the conceptualization of geographic space in thematic regions reflects the way in which spatial behavior is constrained by the regions. If this is the case, that is, if regions are defined by what you can do and cannot do there, we will say that the *behavioral specificity assumption* holds for the spatio-thematic regions in the application domain considered. A queuing area in a ski resort, for instance, permits only a limited number of activities such as entering the queue, moving forward in the queue, and leaving the queue. Other activities, e.g. skiing, are not possible. The behavioral specificity assumption does not claim that we mentally categorize space into regions on the basis of region-specific behavior, although it is likely that we do so in a number of cases. It is an assumption about the application domain which holds whenever the system designer is able to identify for each spatio-thematic region a set of user intentions that are (1) specific to the region

or the region type and (2) characterized by the user's motion patterns. The success of activity-oriented geographic ontologies as a modeling approach for geo-information processing (Jordan et al., 1998; Kuhn, 2001) suggests that both conditions will be met in many application domains.

A basic problem with the interpretation of a motion pattern consists in identifying the subsequences of the pattern that are produced by a specific intention. This *motion segmentation problem* is well-known from cognitively-motivated research on qualitative description of motions (e.g. Musto & al., 2000). A standard approach consists in looking at how properties such as speed, direction, or oscillation of the trajectory change over time. Points where several of the properties change simultaneously are good candidates for the beginning or ending of a segment (Fig. 4). However, segmentation results are often unsatisfactory because the motion pattern by itself contains too few cues to identify segments. Especially, no information about the spatial context of the behavior is considered. In an environment that is divided into spatio-thematic regions, the most salient property of spatial context is its partonomic structure. According to the behavioral specificity assumption, a segmentation strategy driven by the partonomy of the spatio-thematic regions is likely to be successful since the range of possible user behaviors (and intentions) is delimited by regions. Therefore, we propose using a *partonomic segmentation strategy* which chooses the points where the trajectory enters or leaves a region as candidates for the beginning or ending of a segment (Fig. 4).

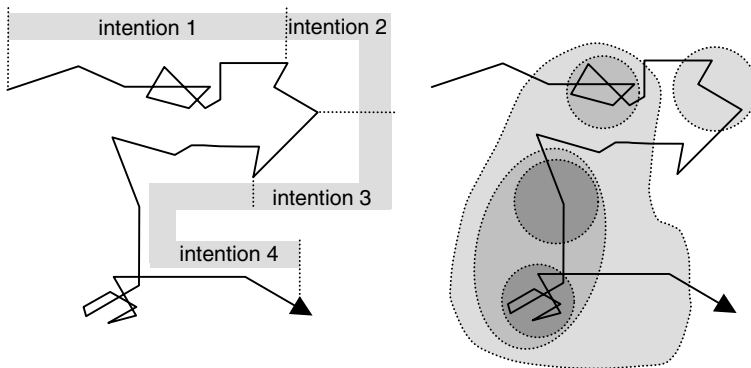


Fig. 4. Segmentation problem and partonomic segmentation

Behavior-Intention Mapping

The number of region-specific user intentions which needs to be distinguished in a motion-based approach is typically very small, that is, rather of order of magnitude 10 than 100. Consider again the museum example. Each exhibit defines a spatio-thematic region within which the visitor must be located in order to study the exhibit more closely. Being located in the region is a necessary but not a sufficient condition for interpreting the visitor's intention as that of studying the exhibit. Additionally, it is required that he stays for a minimum amount of time in the region and that during that time he is oriented towards the exhibit (Tab. 1). In our simplified scenario, we

will not need to distinguish any other type of intention relating to the spatio-thematic region of an exhibit. Note that intentions are not specific to a single region but to a class of regions such as the class of all exhibits.

At higher levels of the partonomy, intentions and the corresponding motion patterns can get quite complex. It requires considerable domain knowledge to describe the spatio-temporal characteristics of these patterns. Information services for museums described by Gabrielli & al. (1999) and Oppermann & Specht (2000) have even drawn on expertise from ethnographic studies. In extensive empirical investigations, Veron & Levasseur (1991) identified four patterns of visitor behavior in exhibitions to which they gave telling names (ant, fish, grasshopper, and butterfly visitors). The classification is based on the objects visited, the time taken, as well as on certain properties of the trajectory. A grasshopper visitor, for instance, pursues a selective non-sequential visit whereas ant-like behavior consists in a complete sequential visit of the exhibition. Tab 1. shows some typical behavior patterns which may be distinguished at the different levels of the partonomy together with associated intentional behaviors.

In order to obtain rules mapping behavior onto intentions, the natural language descriptions of motion patterns that appear in the first column of Tab. 1 have to be expressed in an adequate representational framework. We describe this framework in the following section.

Table 1. Example motion patterns, intentional behaviors and service layers

Motion pattern	Intentional behavior	Level in partonomy	Information service
any type of behavior	visiting	museum	general information
complete sequential visit no crossing behavior at room level	ant-like-touring	wing	map with themes of rooms
selective non-sequential visit crossing behavior at room level	grasshopper-like_touring		map with location of high-lights
walking fast from door to door no studying behavior at exhibit level	crossing	room	themes of neighboring rooms
spending some time in the room studying behavior at exhibit level	visiting		room background information
standing in front of exhibit with orientation towards exhibit	studying	exhibit	information about the exhibit

5 Analyzing Intentions with Partitioned Motion Patterns

The encoding scheme for motion patterns has to meet several requirements. First, there is the need for an adequate representation of the temporal dimension. Second, the encoding should be domain-independent, which implies that it should abstract

from specific sensors. Third, it should be sufficiently expressive to deal with spatial partonomies. In the following, we propose such an encoding scheme.

Encoding Motion Patterns

Formally, a *motion pattern* is defined as a non-empty sequence of *elementary motions* each of which is a 5-tuple of spatio-temporal parameters:

(position, heading, direction, distance, duration).

Position and heading describe the outcome of the motion, that is, the current location of the user and the direction it heads towards. The other parameters give some information about the motion itself. Direction, distance, and duration of the motion are measured with respect to the previous elementary motion (see Fig. 5). Note that the parameters convey redundant information only if they are computed from complete and correct sensor data – an assumption which is rarely given in practice.

Each parameter is specified by a magnitude and a measuring system, e.g. 15,3 m or 3 s. This way, not just the results of quantitative measurements but also those of *qualitative measures*, that is, qualitative abstractions from quantitative measurements, can be stated. Different systems of qualitative spatial and temporal measures have been studied in the field of Qualitative Spatial Reasoning (see Cohn, 1997 for an overview). Typically, magnitudes of qualitative measurements are elements of relational algebras that axiomatize simple computational operations such as relational composition (Ladkin & Maddux, 1994). For instance, a relational algebra defined over {north, west, south, east} allows to express qualitative directions, whereas a relational algebra defined over {near, medium, far} describes qualitative distances.

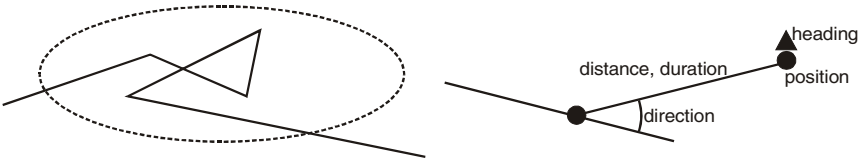


Fig. 5. Motion pattern and parameters of an elementary motion

The 5-tuple of measurement systems of an elementary motion is called the motion's *signature*. Below are the signatures of a quantitative and a qualitative description of a motion.

position: [Gauss-Krüger]	position: {inside, outside}
heading: [radian]	heading: {any}
distance: [meter]	distance: {any}
direction: [radian]	direction: {any}
duration: [second]	duration: {short, medium, long}

The quantitative description is typical for GPS-based localization as it is used, for instance, in the Tourserv project. The qualitative description shown abstracts from all

distance, direction, and heading information. It only indicates whether the agent's position after the motion falls inside or outside the region considered. This matches with region-based sensors which can only detect that the user enters or leaves a region. Obviously, the encoding scheme is sufficiently flexible to handle both, quantitative and quantitative descriptions. Therefore, it fulfils the requirement of sensor-independence. Hybrid descriptions can also be represented. This is a useful feature since many sensors deliver data that is best described by a combination of qualitative spatial parameters (position, heading, distance, direction) and a quantitative temporal parameter (duration).

Behavior rules which interpret the user's spatial behavior as intention sequences are defined with respect to motion patterns. An important task for the information system's designer consists in taking care that the complexity of the behavioral rules matches with the quality of the data delivered by the sensors. Obviously, a behavior rule stating that a visitor shows studying behavior in front of an exhibit if he is closer than 1 m requires that the motion pattern does not completely abstract from distance information.

Motion patterns easily combine with hierarchical data structures that describe spatial partonomies. We represent the way in which a partonomy (or the decomposition tree) divides the motion pattern into subpatterns in the following straightforward way. Each spatio-thematic region delimits a subpattern: the sequence of elementary motions that occur within the region. The organization of the regions in the partonomy is inherited by the subsequences. We call this hierarchical structure a *partitioned motion pattern*.

Analyzing Partitioned Motion Patterns

The primary interest of partitioned motion patterns is that they structure the data about the agent's behavior in a way that reflects the spatial partonomy. To interpret the data, an incremental analysis is run. Each new elementary motion triggers an evaluation cycle. The evaluation tries to associate an intentional behavior to the subpatterns that correspond to *open spatio-thematic regions*, that is, to the regions in the partonomy which contain the agent's current location. Fig. 6 illustrates the notion of open spatio-thematic region. The j -th elementary motion in the i -th subpattern is denoted by m_{ij} . Regions delimiting subpatterns are rectangles that carry the same subscript as the subpattern. The *open subpatterns* defined by the open spatio-thematic regions are shown as open rectangles (r_1 , r_3 , and r_k).

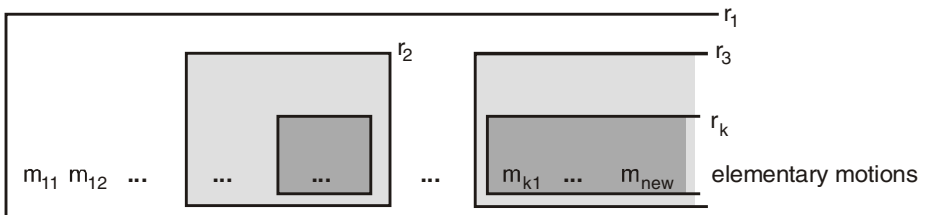


Fig. 6 Identifying intentional behavior from partitioned motion patterns

The interpretation process starts with the open region that has the lowest position in the partonomy (r_k in Fig. 6). Then, the analysis proceeds to the superregions in the order they appear in the partonomy (r_3, r_1). At each level, the behavior rules associated with the spatio-thematic region considered are applied. As soon as a rule fires, a spatio-thematic region has been found in which the motion pattern can be interpreted as an intentional behavior. This most specific behavior is considered the relevant intentional behavior which needs to be supported by information services. As a side effect, the spatial context problem is solved: the first spatio-thematic region with an intentional behavior constitutes the spatial focus of the user's spatial behavior.

For the purpose of illustration consider again the modeling of spatial behavior in the museum domain as specified by Tab. 1. The visitor is located in room n after having spend some time in room i (Fig. 7). At this point the interpretation process starts with interpreting the motion pattern with respect to region n. Interpretation rules at the level of exhibits cannot apply, so the two interpretation rules at the room level are tried. We assume that the behavior is to unspecific to support an interpretation as either visiting or crossing behavior at the room level. In this case, the interpretation rule at the next level, the wing level is tried. It successfully interprets the user's behavior as ant-like-touring behavior. This interpretation is then taken to find supporting information services using the second rule set mapping intentions onto services.

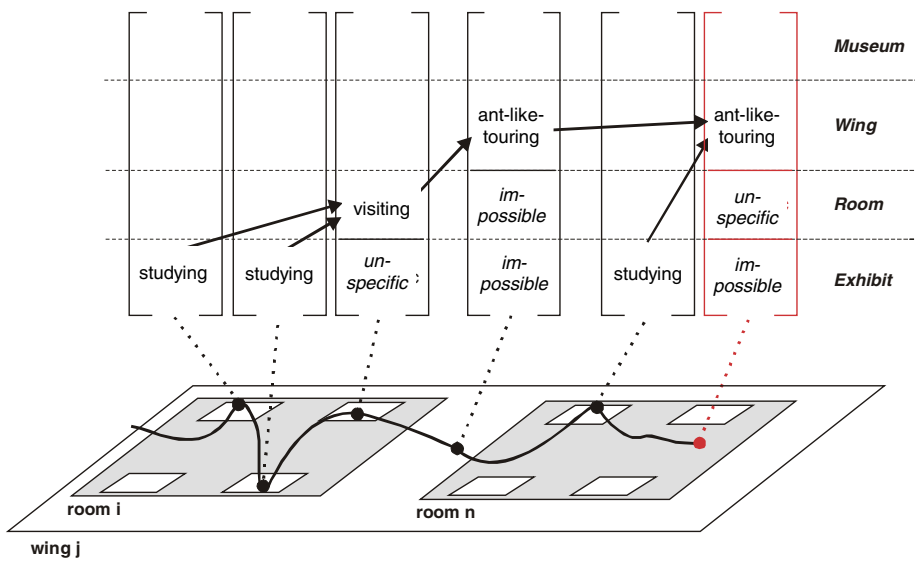


Fig. 7.

6 Related Approaches and Discussion

In a wide range of published work, location is the dominant context parameter to tailor information presentation. Several guide systems have been built (e.g.

Abowd & al., 1997, Davies & al., 1998, Oppermann & al., 2000.) which use the user's current location and his travel history to predict objects of interest to visit. But most of these simply use spatial regions that are closest by, or represent the smallest region for a specific location and do not consider that this location belongs to several spatial regions of a partonomy. If the user is located within the region of a specific object as defined for example by an Active Badge (Want & al., 1992) sensor, the system would decide that this region is the most relevant one and prompt the user with detailed information about this object. We are not aware of any work proposing a cognitively more plausible solution to the spatial context problem that arises from intentional behavior in spatial partonomies.

Mental representations of motions have been studied by researchers in spatial cognition, especially Musto & al. (2000). Based on data from psychological experiments, they propose a qualitative motion representation which uses sequences of qualitative motion vectors. These can easily be expressed in our more general framework as they encode only the direction and distance parameters of the elementary motion we defined. The central concern of Musto & al. (2000) is with a cognitively plausible segmentation of a motion pattern into subpatterns. In our case, however, segmentation is not internal but external, that is, induced by the regions of the partonomy.

In our paper, we have shown that a context problem arises in connection with the interpretation of the user's intentional behavior in a spatial partonomy. We have argued that the observation of the user's motion can provide valuable information for inferring his intentions and we proposed using the partonomic structure of the environment to segment the user's motion pattern into subpatterns which can then be interpreted in terms of intentions. This resolves the spatial context problem as well as the motion segmentation problem. Interestingly, it turned out to be easier to find a solution for both problems than to solve each of them independently which indicates that they are closely interrelated.

The simple representational scheme for encoding motion patterns which we described is sufficiently general to encompass quantitative, qualitative, and hybrid representations of elementary motion. We expect hybrid representations to be especially valuable for the designer who is specifying the rules that describe how to map behavior onto intentions. Finding an intentional behavior for some part of the motion pattern amounts to solving a classification problem. Different algorithmic solutions for classification are available such as neural networks or decision rules. We chose a rule-based approach because it enables the software developer to explicitly state which motion patterns are associated with a specific intentional behavior in the application domain he is modeling. A particularity of the solution proposed consists in introducing an additional modeling layer for the user's intentions. This leads to a two-rule set approach with one set of rules mapping behavior onto intentions and another set of rules mapping intentions onto services.

References

1. Abowd, G., Atkeson, G., Hong, J., Long, S., Kooper, R., and Pinkerton, M. (1997). Cyberguide: A mobile context-aware tour guide. *ACM Wireless Networks*, Vol. 3, pp. 421-433.
2. Cohn, A. (1997). Qualitative spatial representation and reasoning techniques. In *Proc. KI-97: Advances in Artificial Intelligence*, (pp. 1-30). Springer: Berlin.
3. Davis, E. (1990). *Representations of commonsense knowledge*. Morgan Kaufman: San Mateo, CA.
4. Davies, N., Mitchell, K., Cheverst, K., and Blair, G. (1998). Developing a context sensitive tourist guide. In *Proc. of the First Workshop on Human Computer Interaction with Mobile Devices* (pp. 64-68). University of Glasgow, UK.
5. Gabrielli, F., Marti, P., and Petroni, L. (1999). The environment as interface. In M. Caenepeel, and D. Benyon (eds.), *Proc. of the i3 Annual Conference: Community of the Future*, October 20-22, Siena.
6. Hirtle, S. (1995). Representational structures for cognitive space: Trees, ordered trees and semi-lattices. In: *Spatial Information Theory, COSIT-95*, (pp. 327-340), Springer: Berlin.
7. Jordan, T., Raubal, M., Gartrell, B., and Egenhofer, M., (1998). An affordance-based model of place in GIS. In: Chrisman and Poiker (eds.), *Proc. 8th Int. Symposium on Spatial Data Handling, SDH'98* (pp. 98-109). IUG: Vancouver.
8. Kirste, T., Rieck, A., and Schumann, H. (1997) Die Herausforderungen des Mobile Computing: Die Anwenderperspektive. In: *Agenten, Assistenten, Avatare (AAA'97)*, Darmstadt, Germany.
9. Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, 15 (7), pp. 613-631.
10. Ladkin, P., and Maddux, R. (1994). On binary constraint problems, *Journal of the ACM*, 41, pp. 435-469.
11. Musto, A., Stein, K., Eisenkolb, A., Röfer, T., Brauer, W., and Schill, K. (2000). From motion observation to qualitative motion representation. In: Freksa and al. (eds.) *Spatial Cognition II* (pp. 115-126). Springer: Berlin.
12. Opermann, R., and Specht, M. (2000). A context-sensitive nomadic information system as an exhibition guide. *Proc. of the Second International Symposium on Handheld and Ubiquitous Computing*, Bristol.
13. Schilit, B., Theimer, M., and Welch, B. (1993). In: *Proc. of the USENIX Mobile and Location-independent Computing Symposium* (pp. 129-138). Cambridge, MA.
14. Schlieder, C., Vögele, T., and Visser, U. (2001). Qualitative spatial representation for information retrieval by spatial gazetteers. In: *Spatial Information Theory, COSIT-01*, (pp. 336-351). Springer: Berlin.
15. Schlieder (1996). Qualitative shape representation. In A. Frank (ed.), *Spatial conceptual models for geographic objects with undetermined boundaries* (123-140). Taylor & Francis: London.
16. Veron, E., and Levasseur, M. (1991). *Ethnographie de l'exposition: L'espace, le corps et le sens*. Centre Georges Pompidou Bibliothèque Publique d'Information: Paris.
17. Want, R., Hopper, A., Falco, V., and Gibbons, J. (1992). The active badge location system. *ACM Transaction on Information Systems*, Vol. 10, No. 1, pp. 91-102.
18. Winston, M., Chaffin, R. and Herrmann, D. (1987). A taxonomy of part-whole relations. *Cognitive Science*, 11: 417-444.

Author Index

- Albert, William S. 127
- Baier, Volker 305
- Balbani, Philippe 348
- Bateman, John 263
- Blümke, Matthias 209
- Brauer, Wilfried 305
- Cohn, Anthony G. 232
- Condotta, Jean-François 348
- Eschenbach, Carola 89
- Fischer, Kerstin 263
- Freksa, Christian 385
- Galata, Aphrodite 232
- Gattis, Merideth 249
- Habel, Christopher 11, 89
- Haun, Daniel 209
- Hazarika, Shyamanta M. 232
- Höll, Doris 143
- Hogg, David C. 232
- Hommel, Bernhard 157
- Johnson-Laird, P.N. 372
- Klippel, Alexander 11
- Knauff, Markus 372
- Knuf, Lothar 157
- Krieg-Brückner, Bernd 34
- Kuhn, Werner 77
- Kulik, Lars 89
- Lankenau, Axel 34
- Leplow, Bernd 143
- Ligozat, Gérard 348
- Long, Paul 112
- Magee, Derek R. 232
- Mallot, Hanspeter A. 62
- McNamara, Timothy P. 174
- Mehdorn, Maximilian 143
- Mochnatzki, Horst F. 62
- Moratz, Reinhard 263, 385
- Nebel, Bernhard 385
- Pederson, Eric 287
- Rasch, Björn 209
- Röfer, Thomas 34
- Röhrbein, Florian 305
- Schill, Kerstin 305
- Schlieder, Christoph 401
- Schmidtke, Hedda R. 89
- Schönfeld, Robby 143
- Steck, Sibylle D. 62
- Stein, Klaus 305
- Schweizer, Karin 192
- Tappe, Heike 11
- Tenbrink, Thora 263
- Thornton, Ian M. 127
- Timpf, Sabine 77
- Tschander, Ladina B. 89
- Tversky, Barbara 1
- Vorwerg, Constanze 321
- Wender, Karl F. 209
- Werner, Anke 401
- Werner, Steffen 112
- Zetsche, Christoph 305