Christoph Hölscher Thomas F. Shipley
Marta Olivetti Belardinelli John A. Bateman
Nora S. Newcombe (Eds.)

LNAI 6222

Spatial Cognition VII

International Conference Spatial Cognition 2010
Mt. Hood/Portland, OR, USA, August 2010
Proceedings

Springer

Lecture Notes in Artificial Intelligence 6222

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Christoph Hölscher Thomas F. Shipley
Marta Olivetti Belardinelli
John A. Bateman Nora S. Newcombe (Eds.)

Spatial Cognition VII

International Conference Spatial Cognition 2010
Mt. Hood/Portland, OR, USA, August 15-19, 2010
Proceedings

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Christoph Hölscher
Albert-Ludwigs-Universität Freiburg, Center for Cognitive Science
Institute of Computer Science and Social Research
Friedrichstraße 50, 79098 Freiburg, Germany
E-mail: hoelsch@cognition.uni-freiburg.de

Thomas F. Shipley
Temple University, Department of Psychology
Weiss Hall 1701 North 13th Street, Philadelphia, PA 19122-6085, USA
E-mail: tshipley@temple.edu

Marta Olivetti Belardinelli
'Sapienza' University of Rome, Department of Psychology
Via dei Marsi 78, 00185 Rome, Italy
E-mail: marta.olivetti@gmail.com

John A. Bateman
University of Bremen, FB 10, Faculty of Linguistics and Literary Sciences
Building GW2, Bibliothekstraße 1, 28334 Bremen, Germany
E-mail: bateman@uni-bremen.de

Nora S. Newcombe
Temple University, Department of Psychology
Weiss Hall 1701 North 13th Street, Philadelphia, PA 19122-6085, USA
E-mail: newcombe@temple.edu

Library of Congress Control Number: 2010931165

CR Subject Classification (1998): I.2, H.2.8, I.2.10, H.3.1, H.4-5

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-642-14748-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-14748-7 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

This is the seventh volume of a series of books on fundamental research in spatial cognition. As with past volumes, the research presented here spans a broad range of research traditions, for spatial cognition concerns not just the basic spatial behavior of biological and artificial agents, but also the reasoning processes that allow spatial planning across broad spatial and temporal scales. Spatial information is critical for coordinated action and thus agents interacting with objects and moving among objects must be able to perceive spatial relations, learn about these relations, and act on them, or store the information for later use, either by themselves or communicated to others. Research on this problem has included both psychology, which works to understand how humans and other mobile organisms solve these problems, and computer science, which considers the nature of the information available in the world and a formal consideration of how these problems might be solved. Research on human spatial cognition also involves the application of representations and processes that may have evolved to handle object and location information to reasoning about higher-order problems, such as displaying non-spatial information in diagrams. Thus, work in spatial cognition extends beyond psychology and computer science into many disciplines including geography and education. The Spatial Cognition conference offers one of the few forums for consideration of the issues spanning this broad academic range.

This volume represents the fruit of a maturing collaboration between two spatial cognition research centers, one that has emphasized psychology (SILC – Spatial Intelligence and Learning Center, National Science Foundation) and one that has emphasized the computational approach (SFB/TR8 Spatial Cognition, German Research Council DFG). This collaboration began in 2008 with joint work on the sixth Spatial Cognition conference that was held in Freiburg, preceded by an NSF-funded workshop. A smaller, more focused, workshop followed in 2009 in New York City, and collaborative work began on a number of individual projects. The emerging consensus is that spatial cognition research must consider both issues of reasoning about small-scale spatial relations of manipulable objects and the larger-scale spatial problems of navigating among locations. The papers and Keynote speakers of this year's conference reflect this emerging approach and, in the present volume, you will see the results of sustained growth in the field of spatial cognition. The combined efforts of the two centers are contributing significantly to providing the infrastructure necessary for a fully-fledged science of space.

For this volume, 35 papers were submitted and reviewed by at least three members of our Program Committee. Twenty-five papers were selected for presentation and inclusion here. In addition to the submitted papers, the Program Chairs invited three scholars to give keynote lectures. Francesca Pazzaglia, of the University of Padova, Italy, gave a lecture considering individual difference in large-scale spatial thinking entitled "Individual Differences in Spatial Language and Wayfinding: The Role of Cognition, Emotion and Motivation," Kenneth Forbus, of Northwestern University, USA, gave a lecture highlighting work on the computer science approach to education and spatial thinking entitled "CogSketch: Sketch Understanding for Cognitive Science Research and for Education," and Roger Downs, of Pennsylvania State University, USA, gave a lecture linking geography to navigation entitled "The Refraction of Space: A Radical Reversal of Direction." Abstracts of the keynote talks are presented in this volume.

Spatial Cognition 2010 took place at the Resort on the Mountain near Mount Hood Oregon—the first time this conference has been held in North America. In addition to the papers that were presented, nearly 50 posters displayed work in progress. The conference also featured two tutorials, four workshops, and a doctoral colloquium where more than a dozen young scholars had the opportunity to present their research. The Spatial Cognition conference was attended by more than 100 delegates from around the world, including the United States, Germany, Canada, Italy, the United Kingdom, Ireland, Bulgaria and Japan.

Many people contributed to the success of Spatial Cognition 2010. We wish to thank: Adrienne Larmett for the organization and logistics for the conference, David Rapp for his work in organizing the poster sessions, Thomas Barkowsky and Ken Forbus for chairing the Workshop Committee, Kai-Florian Richter for organizing the tutorials, and Andrea Frick, Daniele Nardi, and Kristin Ratliff for organizing the doctoral colloquium. A special thank you goes to Andreas Klein for his help with handling the processing of paper contributions for this volume; and we thank the support staff from SILC and the SFB/TR 8 who helped with the event on site in Oregon.

Finally, we thank Alfred Hofmann and his staff at Springer for their continuing support of our book series.

August 2010

Thomas F. Shipley
Christoph Hölscher
Marta Olivetti Belardinelli
John Bateman
Nora S. Newcombe

Conference Organization

Program Chairs

Thomas F. Shipley
Christoph Hölscher
John Bateman
Nora S. Newcombe
Marta Olivetti Belardinelli

Local Organization

Adrienne Larmett

Tutorial Chair

Kai-Florian Richter

Workshop Chairs

Kenneth D. Forbus
Thomas Barkowsky

Poster Session Chair

David Rapp

Doctoral Colloquium Chairs

Daniele Nardi
Kristin Ratliff
Andrea Frick

Program Committee

Pragya Agarwal
Marios Avraamides
Thomas Barkowsky
John Bateman
Michela Bertolotto
Stefano Borgo
Melissa Bowerman
Angela Brunstein
Wolfram Burgard
Laura Carlson
Anjan Chatterjee
Christophe Claramunt
Eliseo Clementini
Anthony Cohn
Ruth Conroy-Dalton
Leila De Floriani
Franco Delogu
Maureen Donnelly
Matt Duckham
Russell Epstein
Kenneth D. Forbus
Christian Freksa
Antony Galton
Dedre Gentner
Gabriela Goldschmidt
Louis Gomez
Christian Graf
Klaus Gramann

Glenn Gunzelmann
Christopher Habel
Mary Hegarty
Kathy Hirsh-Pasek
Stephen Hirtle
Christoph Hölscher
Tina Iachini
Gabriele Janzen
Angie Johnson
Jonathan Kelly
Alexander Klippel
Markus Knauff
Maria Kozhevnikov
Antonio Krueger
Yohei Kurata
Unmesh Kurup
Gerhard Lakemeyer
Longin Jan Latecki
Susan Levine
(Lily) Chao Li
Hanspeter Mallot
Justin Matthews
Mark May
Timothy P. McNamara
Tobias Meilinger
Daniel R. Montello
Stefan Muenzer
Nora Newcombe
Marta Olivetti Belardinelli
Dimitris Papadias
Eric Pederson
Fiora Pirri
Markus Plank
Ian Pratt-Hartmann
Antonino Raffone
Marco Ragni
Martin Raubal
Terry Regier
Kai-Florian Richter
Andrea Rodriguez Tastets
Valerio Santangelo
Kerstin Schill
Ute Schmid
Werner Schneider
Holger Schultheis
Kathleen Stewart
Holly Taylor
Thora Tenbrink
Barbara Tversky
Florian Twaroch
David Uttal
Nico Van de Weghe
Constanze Vorwerg
Stefan Woelfl
Thomas Wolbers
Diedrich Wolter
Wai Yeap

Table of Contents

Science Education and Spatial Skill

Language

Computational Modelling

Reference Frames

Individual Differences in Spatial Language and Way-Finding: The Role of Cognition, Emotion and Motivation

(Abstract)

Francesca Pazzaglia and Chiara Meneghetti

University of Padua, Department of General Psychology (DGP), Via Venezia 8, 35131 Padua, Italy

People can experience an environment in different ways and from different points of view: by moving around in it, inspecting it from above (flying, viewing from a mountain top or high building), studying a map, or listening to a verbal description. How an environment is experienced can influence its spatial representation and, as a consequence, spatial performance. However, this latter can also be influenced by a series of factors inherent to individuals: gender, spatial and working memory abilities, cognitive styles in spatial representation, motivation and attitude toward spatial tasks, emotion and personality. Here I present an overview of a series of research programs underway in the Laboratory of Learning and Memory at the University of Paduas Department of General Psychology. Our research on spatial cognition has focused on two main topics: (i) memory and comprehension of spatial language, (ii) navigation and way-finding behavior. In both cases, particular emphasis has been given to the study of individual differences. Spatial text comprehension and navigation vary widely among individuals. We investigated potential sources of individual differences and examined the roles of spatial ability, working memory, cognitive style in spatial representation, as well as those of other non-cognitive variables: motivation, personality and emotion. To measure individual differences in sense of direction (SOD), cognitive styles in spatial representation, motivation in performing spatial tasks, and spatial anxiety, we constructed a number of self-rate scales, which are illustrated in terms of performance on various spatial tasks.

Individual Differences in Spatial Text Comprehension

First I shall present the main results of a series of studies on the role of spatial ability, working memory, and imagery in the comprehension of spatial descriptions (De Beni, Pazzaglia, Gyselinck, & Meneghetti, 2005; Gyselinck, Meneghetti, De Beni & Pazzaglia, 2009; Pazzaglia, De Beni & Meneghetti, 2007). In this context, people construct a mental (or situation) model that maintains the spatial features of the environment described (Perrig & Kintsch, 1985; Taylor & Tversky, 1992). In doing so, both verbal and spatial components of working memory are involved.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 1–3, 2010.

To what extent do limitations in spatial working memory and spatial ability influence comprehension of spatial texts? Are imagery strategies efficient in promoting spatial text comprehension and to what extent do they need visuo-spatial working memory (VSWM) to be implemented? Finally, do cognitive styles in spatial representation (i.e. individual tendency to represent environment from survey or route perspective) influence the comprehension of spatial texts? Overall, the study results I will present support the notion that VSWM is strongly implicated in spatial text processing, but also that individual differences in spatial ability (e.g. mental rotation) and spatial representation can explain performance in spatial text comprehension.

Individual Differences in Way-Finding

In the second part of my presentation I shall review our groups studies on individual differences in way-finding, studied in both real and virtual environments. The role of preference for route or survey representation is explored. By presenting the key results from a number of studies (Denis, Pazzaglia, Cornoldi & Bertolo, 1999; Pazzaglia, Taylor, 2007) I shall demonstrate that individuals differing in cognitive style in spatial representation (high and low survey individuals) perform differently in way-finding tasks, and that cognitive style interacts with spatial representation and instructions in influencing performance. Finally I shall present the main findings from one of our research project investigating whether personality traits, motivation toward orienting tasks, spatial anxiety, in addition to spatial ability and VSWM, can explain performance on two distinct tasks: navigation through a just-learned route, and finding a short-cut. In this study a sample of 115 undergraduates were given: (1) a battery of tests for the assessment of spatial ability and VSWM; (2) self-rate scales on SOD, spatial representation, motivation, and spatial anxiety; (3) the BIG Five Questionnaire (Caprara, Barbaranelli & Borgogni, 2000) for assessment of personality traits. A structural equation model (SEM) was computed using the LISREL 8.7 statistical package (Jreskog & Srbom, 1996). The outcomes show that the considered variables are involved in various different ways in explaining the performance of the two orienting tasks: cognitive measures (spatial ability and VSWM) predict navigation through an experienced route; instead, short-cut finding involves non-cognitive variables personality, emotion and motivation.

Conclusions

Spatial text comprehension and orientation are complex tasks that involve problem-solving processes. It follows that a variety of cognitive and non-cognitive factors should be taken into account - in addition to their interconnections - when seeking explanation for this complexity. Differences due to materials, instructions and task type also need to be considered. The overview given in this present paper highlights the need to develop an integrated model of spatial cognition, able to provide a clearer explanation of individual behavior and to support implementation of more efficient aids for orientation, tailored to individual profiles.

References

1. Caprara, G.V., Barbaranelli, C., Borgogni, L.: Big Five Questionnaire. Organizzazioni Speciali, Florence, Italy (2000)
2. De Beni, R., Pazzaglia, F., Gyselinck, V., Meneghetti, C.: Visuo-spatial working memory and mental representation of spatial descriptions. European Journal of Cognitive Psychology 17, 77–95 (2005)
3. Denis, M., Pazzaglia, F., Cornoldi, C., Bertolo, L.: Spatial discourse and navigation: An analysis of route directions in the city of Venice. Applied Cognitive Psychology 13, 145–174 (1999)
4. Gyselinck, V., Meneghetti, C., De Beni, R., Pazzaglia, F.: The role of working memory in spatial text processing: What benefit of imagery strategy and visuospatial abilities? Learning and Individual Differences 19, 12–20 (2009)
5. Jreskog, K.G., Srbom, D.: LISREL 8: Users reference guide. Scientific Software International, Chicago (1996)
6. Pazzaglia, F., De Beni, R., Meneghetti, C.: The effects of verbal and spatial interference in the encoding and retrieval of spatial and nonspatial texts. Psychological Research 71(4), 484–494 (2007)
7. Pazzaglia, F., Taylor, H.A.: Perspective, Instruction, and Cognitive Style in Spatial Representation of a Virtual Environment. Spatial Cognition & Computation 7(4), 349–364 (2007)
8. Perrig, W., Kintsch, W.: Prepositional and situational representations of text. Journal of Memory and Language 24, 503–518 (1985)
9. Taylor, H.A., Tversky, B.: Spatial mental models derived from survey and route descriptions. Journal of Memory and Language 31, 261–292 (1992)

CogSketch: Sketch Understanding for Cognitive Science Research and for Education
(Abstract)

Kenneth D. Forbus

EECS Department, Northwestern University, Ford 3-320, 2133 Sheridan Road, Evanston, IL, 60208, USA

Sketching is a powerful means of working out and communicating ideas. Sketch understanding involves a combination of visual, spatial, and conceptual knowledge and reasoning, which makes it both challenging and potentially illuminating. This talk will describe how a team of AI researchers, cognitive psychologists, learning scientists, and educators is attempting to build the intellectual and software infrastructure needed to achieve more human-like sketch understanding software. We are creating CogSketch, an open-domain sketch understanding system that will serve as both a cognitive science research instrument and as a platform for sketch-based educational software. These missions interact: Our cognitive simulation work leads to improvements which can be exploited in creating educational software, and our prototype efforts to create educational software expose where we need further basic research. CogSketch incorporates a model of visual processing and qualitative spatial representations, facilities for analogical reasoning and learning, and a large common-sense knowledge base. Our vision is that sketch-based intelligent educational software will ultimately be as widely available to students as graphing calculators are today.

I will start by describing the basics of open-domain sketch understanding and how CogSketch works. Some cognitive simulation studies using CogSketch will be described, to illustrate that it can capture aspects of human visual processing. The potential use of implicit, software-gathered measures of expertise for assessment will be discussed, based on a recent experiment with sketching in geoscience. Two prototype educational software efforts will be summarized. The first, worksheets, provides a simple way to see if students understand important configural relationships, e.g., the layers of the Earth. The second, the Design Buddy, is intended to help students learn how to communicate via sketching in the context of learning engineering design.

While CogSketch is a work in progress, the current prototype is publicly available, and we seek community feedback and collaboration. CogSketch can be downloaded at http://www.silccenter.org/initiatives/tools/sketching.html.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, p. 4, 2010.

The Refraction of Space: A Radical Reversal of Direction (Abstract)

Roger M. Downs

Penn State University, 312B Walker Building, University Park, PA 16802, USA

The field of spatial cognition began nearly fifty years ago. In the early years, Kevins Lynch's *The Image of the City* was an inspirational icon for many while Kenneth Boulding's *The Image* and E.C. Tolman's "Cognitive maps in rats and men" provided succinct but elementary philosophical and psychological groundings.

The disciplinary roots of the field were disparate: architecture and landscape architecture, planning, psychology, geography, sociology, economics, and anthropology were the principal sources of scholars. If for the moment we take psychology as a benchmark for comparison, all of the other disciplines had some things in common in their approaches to spatial cognition.

Unlike psychology, they were not experimental in character: they sacrificed the power of systematic manipulation and careful control for the appealing but ad hoc messiness of ecological validity. They were less rigorously scientific than psychology: they freely mixed the speculative hand-waving associated with artistic and humanistic descriptions with the explanatory requirements of scientific precision. They were pragmatic in orientation: they often traded off the long-term development of abstract theory for the immediate benefits of short-term practical applications to real world situations.

While these methodological differences were, and perhaps still are, significant, the major conceptual distinction is in the contrasting emphases placed on the role of the adjective, spatial, versus that of the noun, cognition. For the other disciplines, the preferred adjective was as often environmental as it was spatial. Whichever term was used, however, the idea of space or environment was the focus.

For psychology, the focus, and thus the starting point, has been the noun, cognition. This position was based a recognition that the cognition of space or the environment was more than just a special case of applying general cognitive principles to one of many possible substantive domains. The rats which refused to be bound by the confines of the maze walls were taking physical short cuts *and* making cognitive leaps that required new explanatory models of thinking.

For the other disciplines, the focus and starting point has been the attempt to explain and therefore predict spatial behavior in the environment. The field of spatial cognition was a means to understand what happened in terms of behavior in the environment and equally well, a vehicle to effect behaviorally beneficial changes in the design of the environment. Kevin Lynch's links between a city's

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 5–6, 2010.

imageability and therefore legibility provided design tools for reshaping the built environment of the city.

I recognize that this distinction is perhaps over-simplified and thus overstated. In practice, there are certainly not two parallel but only loosely connected streams of research. Equally well, I would and will argue that as the field of spatial cognition has developed, relatively speaking the emphasis has been on the cognition side of the term. While I do not disparage such an approach, I do believe that we can profitably go back to look more carefully at the adjectival modifiers, spatial and environmental.

As a geographer, to me understanding the idea of environment requires an explicit combination of two fundamental and interrelated concepts: space in the abstract and place in the particular. As a relational framework, space is neutral and therefore value-free. Place, on the other hand, is loaded with layers of symbolic meanings. In this perspective, work in spatial cognition should reflect a creative tension between the understandings of space and place.

In words that will resound with Trekkies, I believe that space remains the final frontier. In this talk, I want to explore, although perhaps not boldly, the ideas of space and place as they relate to spatial cognition. The radical part of my title refers to the literal idea of roots and it is to the roots of spatial models that I want to return. The reversal part of my title is an appeal to refocus our attention on the twin concepts of space and place.

I want to use the mathematical properties of metric spaces to understand how the field of spatial cognition has dealt with space and place. In doing so, I will use some of the classics in the field such as Segall, Campbell, and Herskovits (1966), Stevens and Coupe (1978), and Tversky (1992) to unpack the properties, especially the implicit properties, of models of space and place. In line with my training as a geographer, I will draw heavily on the relations among spaces and mappings to understand how the choice of spatial models shapes our understanding of spatial cognition and spatial behavior. I will argue that we have perhaps over-emphasized the cognition part of the title at the expense of the spatial part of the title and that we have tended to lose sight of the spatial behaviors that originally motivated our interest.

Investigating the Role of Goals and Environmental Structure on Memory for Distance and Time in Virtual Environments

Angie Johnson[1], Kenny R. Coventry[1], and Emine M. Thompson[2]

[1] Cognition and Communication Research Centre, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
[2] Built Environment Visualisation Centre, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
{angie.johnson,kenny.coventry,emine.thompson}@northumbria.ac.uk

Abstract. Individuals rarely walk in an environment without a purpose. However, the influence of goals on the development of 'cognitive maps', has largely been ignored. The results of two experiments are reported that investigated the role of both goals and environmental structures on memory for distance and time in Virtual Reality (VR) environments. Experiment 1 compared the effect of goals varying in urgency and desirability on memory for distance and time in VR environments with and without turns. Striking effects of goals were found for memory for distance and time in both environments. Experiment 2 examined the origins of these goal effects through the use of physiological measurement and mood scales. Results show that goals influence distance estimation as a function of the degree of urgency experienced in situ, and not as a function of overall mood state or arousal on the estimation process.

Keywords: Cognitive Maps, Virtual Reality, Goals, distance and time perception.

1 Introduction

The development of 'cognitive maps' enable individuals to navigate through environments from memory [1, 2]. There is evidence that they can be created from actually walking in an environment or, through exposure to a variety of media, such as maps [3], diagrams [4], photographs, and verbal descriptions [5, 6]. The use of Virtual technologies in the study of how spatial information is encoded and retrieved has been a relatively recently addition to the spatial cognition researcher's toolkit. However, whilst a range of secondary methods facilitate the construction of cognitive maps [7, 8] the issue of whether their effectiveness can be functionally equivalent to the primary experience of actually walking in space has received much attention.

It is well known that memory for spatial environments, whether they are learned from moving around real space or from other sources (as in VR), is associated with a range of perceptual distortions. These can be due to environmental features, such as barriers [9], the function and familiarity of landmarks [10], or the number of turns in an environment [11], all affecting memory for distance. In addition, there is evidence

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 7–18, 2010.

suggesting that individual differences play a part in inconsistencies between actual and perceived distance and time [12], including the effort taken to move around the environment [13], and goals [14, 15]. Therefore, experience of an environment is not wholly dependent on either environmental features or visual information, as individual differences appear to mediate these influences.

One such individual variable that may critical to the acquisition and memory of spatial information in our environment is that of goals. People rarely walk in an environment without a reason – from posting a letter to buying groceries – but studies in spatial cognition seldom involve ecologically valid goals, except in those studies where participants simply have to learn a new environment or wayfind to a new place. Moreover, thus far, comparisons between spatial learning in real versus virtual environments have as yet only just begun to examine how goals might affect perception of distance and time in these different media.

Johnson, Coventry & Thompson [16] tested whether goals affect memory for distance and time in two real 'human maze' environments made of large polystyrene cuboids - one involving a straight path, and one involving multiple turns. Using a role play scenario, participants were requested to walk (in time with a metronome to control speed of walk and time spent walking) to the end of a path to deliver an object to a friend. The scenarios used varied in urgency and desirability, such as delivering medication to a sick friend (high urgency, high desirability) or delivering exam results to a friend, opening the envelope only to find that they had failed everything (low urgency, low desirability). A between participants design ensured that there were no transfer effects concerning the conditions or scenarios, and pre-experiment distance judgments found no significant differences amongst the participants in their ability to estimate distance. Following the delivery of the object, participants in the urgent scenario reported the path previously walked as being significantly longer than the participants in the non-urgent scenario. However, an effect of goals was not found in the real space path with many turns, and several participants reported that they experienced difficulty walking in time to the metronome whilst turning the corners, thus distracting them from the task at hand. This might also explain why a robust 'turns' effect [11] was not found, as Montello [17] has noted, secondary tasks can divert attention away from one's environment.

The main aim of the studies we report below was to test whether goals affect memory for time and/or distance in virtual environments. If the failure to find an affect of goals on memory for distance in a real environment with turns in Johnson *et al.* [16] was due to awkwardness moving around the turns in the real environment, the removal of this awkwardness in VR should produce effects of goals. Additionally, it is the case that behavior in virtual environments, while generally similar to behavior in the real environments, does differ in a number of respects. Actual distances are consistently underestimated in virtual environments compared to their real space alternatives, even when realistically portrayed [18]. Perceptual errors have been observed due to a variety of contributory factors, such as reduced field of view or other constraints such as eye height or optic flow [19]. However, there are areas where it is claimed that VR enhances real space research, such as distance knowledge, acquisition of route and survey knowledge [20, 21]. For VR to be an effective methodology, we hoped that an effect of goals would be present, at least for the straight path, as this was found in the real environment in Johnson *et al.* [16].

A further goal of the present studies was to try to unpack reasons for effects of goals on memory for virtual environments, should they be present. One possibility is that the role play scenarios simply differ in general levels of physiological arousal induced, and/or affect the mood state of the participants about to perform any task. Arousal is generally non specific [22] and can easily be transferred from one source to another, influencing perception, as well as attention and memory [23, 24]. It is possible that role play scenarios simply induce states that affect any kind of estimation process. To test this possibility, it is therefore desirable to get participants after role play induction to estimate distances and times without walking those distances. The second experiment below examines this possibility.

2 Experiment 1

2.1 VR Straight Paths versus VR Path with Turns

This experiment examined the effect of goals, varying in urgency and desirability, on memory for distance and time taken to travel (seconds) in a VR straight line route and a VR path with several turns.

2.2 Methodology

The methodology utilized was exactly the same as that used for the Real Space experiments in Johnson *et al.* [16]. Participants were informed that they were going to take part in a 'role play' experiment in VR. The methodology controlled for potential confounding variables such as visual cues, walking pace, and time spent within the environment. Natural walking pace was calibrated to the sound of a metronome affording participants the sensation of walking comfortably and effortlessly [following 25].

The 'role play' scenarios were also adopted from the Johnson *et al.* [16] study. Two separate scenarios previously identified as high and low in urgency were used, which also corresponded to the predicted speed of walk during such circumstances.

- High Urgency – delivering medication to a friend critically ill in hospital.
- Low Urgency – delivering exam results to a friend, and opening the envelope so that you know that they have failed.
- Control Condition – no scenario.

In order to replicate the Real Space task, paths were marked with tape on the floor, to emphasize the start and end of the task. The requirement that participants walk in time with the metronome in real space eliminated the possibility that participants would simply walk more quickly (or run) through the maze in the more urgent scenario. In VR, the requirement was to ensure that the visual flow corresponded with the sound of the metronome and head bob.

Virtual Reality Environment

The computer model depicted an exact replica of both the real space straight path and path with several turns [16] using 3DStudioMax software (see Figure1). The model was extremely realistic, replicating light and textures that gave the strong impression of the original real space paths. The Virtual Environment used consisted of Intel Xeon X5450 CPU 2 x Quad (8 cores running at 3.00 Ghz) GPU: Nvidia Quadro FX 5600, and StereoWorks DLP Passive Stereo Projection System that was based on a rigid rear projection screen. Dimensions were: 244x183cm, with images projected by two Christie Digital DS+25 high resolution projectors. The VR condition was passive and the pace pre-programmed according to the individual's natural walking speed. A 'head bob' of 1.5cm was used to enhance the perception of natural walking [25].

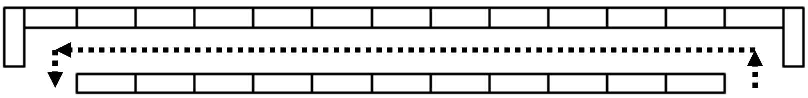


Fig. 1a. Allocentric View – Virtual Replication of Straight Real Space Path made from Polystyrene Blocks (15.4 meters)

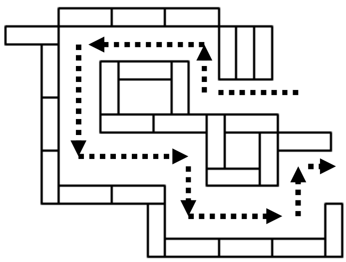


Fig. 1b. Allocentric View – Virtual Replication of Real Space Path with Turns made from Polystyrene Blocks (15.4 meters)

Presentation

The study was presented as a role play in VR task and participants were told that they were going to be transported to the experimenter's 'Virtual Polystyrene Block World' where a scenario would be described to them. Their task was to imagine themselves as the main character in the scenario whilst delivering the object. Great emphasis was placed upon the scenario so that participants would take the task seriously and treat it as being as real as possible. Participants were advised that they would be asked questions about their experiences at the end of the experiment. They were also naïve to the main purpose of the experiment, which investigated their memory for distance and time during the task.

Experimental Design

The experiment employed a 2 (path structure: straight vs. turns) x 3 (scenario: High Urgency vs. Low Urgency vs. Control) between participants design to examine the influence of goals on distance and time estimation.

Participants
One Hundred and Twenty individuals were recruited from the University of Northumbria and were paid a nominal fee or given a course credit (if they were students) for their participation. A wide age range was selected, between 18 and 53 years (mean age 25.4, SD = 8.6), to reflect the diverse population amongst which VR is utilized. All participants were randomly allocated and evenly distributed for age and gender across all of the conditions. No participant had any experience of the constructed VR environment prior to the study.

2.3 Procedure

Participants were tested individually in a session lasting about 30 minutes. Initially they were met in an area separate from the VR environment where they were instructed to walk along a pre-designated walking path, to establish their speed of walk and step length at their natural walking pace. They were also asked about their current level of anxiety; "*Please rate your current level of anxiety on a scale of 1 to 10 – where 1 is 'not anxious at all' and 10 is 'extremely anxious'*"

This was then used as a baseline measurement. Next participants were advised that they were to be blindfolded and transported to the experimenters 'Virtual Polystyrene Block World.' Once in the VR Suite, participants walked a VR reference path, created to the same dimensions as that used earlier in real space. They were exposed to this path three times in order to get used to the VR environment and to establish if the clicks of the metronome, together with the visual flow and head bob, felt natural and reflected their normal walking pace. The participants were then blindfolded and seated in front of the screen to allow for the transition to the main path. Then, participants, standing but still wearing the blindfold, were reminded of the 'role play task' that they had agreed to take part in. When confirmation was received that they were ready to take part, the participant was instructed to visualize the assigned scenario and, following confirmation that they understood what was required, the experimenter gave the participant a medicine bottle or folded exam result manuscript to deliver at the end of the path. Immediately on receipt of the object, the participants were advised to remove their blindfold and commence their journey.

Participants' navigation times were pre-set in accordance with their step length and speed of walk prior to setting up the VR environment. On completion, participants were asked a range of questions concerning distance estimation (in meters) and time estimation by asking:-

a) How long do you think the path you travelled was?
b) How long did you think it took you i.e. length of time, to walk the path?.

They were also asked to report, on a Likert Scale from 1 to 10, how urgent they thought the task was and they were, with 1 being "not at all anxious/urgent" and 10 being "extremely anxious/urgent". Subjective anxiety measurements were taken during and towards the end of the task, with a final subjective report of how anxious they were when the task was over.

2.4 Results

Preliminary Analysis

A series of between participants one-way ANOVAs were run to check that participants across conditions did not differ in their walking speeds or time spent in the VR environments. No significant differences were found for the pace of walk for the straight path, F (2, 52) = 1.81, $p > 0.05$ or the path with several turns, F (2, 52) = 1.54, $p > 0.05$ between scenarios. There were also no significant differences in the times spent in the VR environments between scenarios for the straight path $F(2,52) = 0.11$, $p > 0.05$ or the path with turns $F(2,52) = 1.53$, $p > 0.05$.

Main Analyses

Responses from 4 participants (7%) in the straight path and 5 (8%) in the path with turns were excluded as their distance estimations exceeded the criterion of 2 standard deviations from the mean. Responses from 56 participants in the straight path and 55 from the path with turns were included in the following analyses.

The navigated route, that is, the actual path that participants walked in the path with turns had to be ‘smoothed’ at the corners to create a more naturalistic feel, which also shortened the route for both distance and time. Therefore distance and time estimations were converted to ratios: estimated distance/actual distance in order to provide comparable reports across conditions (see Tables 1 and 2).

Distance Estimation

A 2 (path structure: straight path and the path with 8 right-angled turns) x 3 (scenario: high, low, control) between subjects ANOVA was performed on distance ratio estimations. There was a significant main effect of structure F (1, 110) = 10.99, $p < 0.001$, with the path with turns being reported as significantly longer (M = 1.25) than the straight path (M = 0.79). There was also a main effect of scenario F (2,110) = 3.33, $p < 0.05$, with urgency influencing distance estimation. However, there was no reliable interaction between path structure and scenario. F (2,110) = 1.81, $p > 0.05$. Tukey posthoc tests of the three conditions indicate that the high urgency condition demonstrated a trend approaching significance for longer distance estimations compared to the control condition (M = 0.4), MSe=0.17, p = 0.054. The low urgency condition also produced longer distance estimations compared to the control condition (M = 0.35), MSe = 0.17, p = 0.098. There were no significant different reports of distance estimations between the high urgency and low urgency conditions, (M = 0.05), MSe = 0.17, p > 0.05.

Table 1. illustrates the mean ratio distance (meters) estimates and standard deviations by scenario and path structure

Scenario	Straight Path		Path with Turns	
Distance Ratio Estimated/Actual	$\bar{X}$	(σ)	$\bar{X}$	(σ)
Control	0.69	(0.49)	0.86	(0.51)
High Urgency	0.98	(0.89)	1.35	(0.89)
Low Urgency	0.71	(0.47)	1.52	(0.91)

Time Estimation

There was no significant main effect of path structure F (1, 110) = 0.51, $p > 0.05$ and no reliable interaction between path structure and scenario F (2,110) = 0.115, $p > 0.05$ on time estimation. However, there was a significant main effect of scenario F (2,110) = 6.01, $p < 0.005$. LSD pairwise comparisons of the three conditions indicated that the high urgency scenario (M = 3.9), resulted in significantly longer estimations than the low urgency scenario (M = 1.81), MSe = 0.77, p < 0.01 and also the control condition (M = 1.28), MSe = 0.78, p < 0.001. There were no significant differences between the control and low urgency scenario, p > 0.05.

Table 2. Illustrates the mean ratio time (sec) estimates by scenario and path structure

Scenario	Straight Path		Path with Turns	
Time Ratio Estimated/Actual	$\overline{X}$	(σ)	$\overline{X}$	(σ)
Control	1.13	(0.9)	1.44	(0.71)
High Urgency	3.56	(4.88)	4.1	(5.55)
Low Urgency	1.56	(1.91)	2.05	(2.64)

Urgency

The urgency reports were significant according to scenario F(2, 110) = 32.11, $p < 0.001$, but not for path structure, F(1,110) = 0.00, $p > 0.05$, and there was no interaction between structure and scenario, F(2,110) = 0.186, $p > 0.05$. A significant correlation was found between the level of urgency reported and the time estimations, $r = 0.295$, $p < 0.005$, but there was no significant correlation between the level of urgency reported and distance estimation, $r = 0.151$, $p > 0.05$.

Anxiety

Changes in self reported anxiety measurements were also significant according to scenario F(2,110) = 4.97, $p < 0.0001$, and path structure, F(1,110) = 27.4, $p < 0.001$. There was no reliable interaction between path structure and scenario F(2,110) = 2.3, $p > 0.05$.

2.5 Discussion

The results confirmed the expectation that goals distort distance estimations across path structures and also provided further support for the robust effect of turns on distance estimation found in past studies [11, 25]. It also sheds light upon the failure of Johnson *et al.* [16] to establish the effect of turns in the real space path, supporting the explanation that cognitive resources were directed towards keeping in time with the metronome at the corners thus distracting participants from both the environment and the role play task.

The study also highlights the usefulness of VR as a valid research tool as it can tease apart a variety of factors in very strict and controlled conditions, despite acknowledgements that it does not fully simulate the real life experience [19].

Subjective urgency ratings immediately following the experiment indicated that participants immersed themselves in the role play appropriately, which is also supported by the self-reported anxiety levels. The results from these studies and the original real space straight path study [16] provide evidence that goals influence the immediate memory for distance in both real space and VR across environmental path types. Moreover the effects of goals in the present study were also robust for time estimations. However, the underlying mechanism for these effects now needs to be established. Experiment 2 attempts to begin to decompose the psychological factors contributing to the effect of goals by examining the role of arousal and mood arising from the task instructions.

3 Experiment 2

The effect of goals on memory for distance and time could be due to the scenarios affecting how participants encode space when they are performing the task. However, an alternative explanation is that increasing arousal/anxiety levels associated with the different role play scenarios may influence the estimation process, irrespective of immersion in the task. Indeed there is ample evidence to suggest that arousal from emotions, such as anxiety and excitement, can influence attention, memory and perception [27, 28]. As a result it is necessary to determine whether arousal contributed to the distortion of distance estimation in the previous studies by getting participants to estimate a distance after a scenario induction without walking that distance in the scenario context. If arousal/anxiety increases alone explain distance estimation, an effect of scenario should be present even for a route they have walked prior to the scenario induction.

3.1 Method

The method used was exactly the same as in Experiment 1.

Materials
VR models were exactly the same as for the VR environment in Experiment 1. In addition, a SpOT+pulsoximeter was used to measure heart rate and a 14 item Positive Affect Negative Affect Scale (PANAS) was used to measure change in mood. The PANAS, questionnaire consisted of seven positive and seven negative mood states adopted from the sort version, and has reliably been demonstrated to identify personality states and traits such as anxiety [29].

Participants
Fifty Four participants were recruited to take part from Northumbria University and were paid a nominal fee for their participation. Participants were aged between 18 and 38 years old (mean age 22.6, SD= 3.9). Again, participants were randomly allocated and evenly distributed for age and gender across all conditions. No participant had any previous experience of the room in which the VR environment was located prior to the study.

3.2 Procedure

The procedure was the same as for Experiment 1, with the exception that participants had their heart rates monitored prior to walking the real space reference path as a baseline measurement. Heart rate was also monitored and discretely noted at the beginning of the induction of the scenario and a final heart rate was noted when the participants removed the blindfold in order to deliver the object. Participants did not walk the main VR path; instead they were asked a series of questions; starting with an estimation of the real space reference path they walked, followed by the 14 questions adopted from PANAS and subjective urgency and anxiety ratings for the task.

3.3 Results

Main analyses
Outliers identified as exceeding the criterion of 2 standard deviations from the mean for distance estimations, were excluded. This resulted in 2 participants being excluded from the analyses. In total, 52 participants were included in the following analyses.

Distance estimation

Table 3. Mean distance estimations (m) and standard deviations for the pre-test reference path

Scenario	Control		High Urgency		Low Urgency	
	$\bar{X}$	(σ)	$\bar{X}$	(σ)	$\bar{X}$	(σ)
Distance (actual 8 meters)	6.1	(2.5)	5.9	(2.6)	5	(2.6)

Heartrate
There were no significant changes in heart rate across the conditions throughout the course of the experiment. The first heart rate calculation represents the change in heart rate from baseline until immediately prior to describing the scenario $F(2,49) = 1.08$, $p>0.05$. There was also no significant change in heart rate from heart rate from the induction of the scenario until the time when the participant was advised to remove the blindfold to deliver the object, $F(2,49) = 1.21$, $p > 0.05$.

Urgency and Emotion
The urgency reports were significant according to scenario $F(2, 49) = 16.31$, $p < 0.001$, which is in line with the previous experiments. Significant correlations were found between the level of urgency and the subjective level of negative emotion reported, $r = 0.499$, $p <0.001$. There was also an effect of gender, with males reporting higher positive emotions than females, $r = -0.314$, $p < 0.05$.

In addition, the negative mood scores were reliably different across the scenarios F $(2, 49) = 6.3$, $p < 0.005$, but not the positive mood scores, F $(2, 49) = 0.65$, $p > 0.05$. There were no significant correlations between negative mood and distance estimation, $r = -0.173$, $p > 0.05$, or positive mood scores, $r = 0.126$, $p > 0.05$.

3.4 Discussion

The results show that the effects observed in the experiment 1 were not due to physiological arousal or changes in mood states induced by the scenarios. Most importantly, after the scenario had been induced, participants across conditions did not differ in their distance estimates for the previously walked reference path. Moreover, levels of heart rate change across conditions did not differ between conditions, nor did self-reported mood state. However, self reported urgency ratings confirmed that participants engaged with the role-play task, for instance the both role-play scenarios evoked higher urgency reports than the control condition. This pattern has been consistent across the previous experiments. These outcomes therefore rules out possible contributory factors of either physiological arousal or mood change evoked from the scenarios. The distortion of distance and time estimation is due to other factors involving the actual performance of the task associated with the scenarios.

4 General Discussion

The first experiment investigated whether urgency of goals influences distance and time estimations across environmental structures in virtual environments, following previous studies which confirmed the fidelity of VR as a tool for capturing real environmental experiences [20, 16].

The results of the first experiment revealed a robust effect of turns on memory for distance, consistent with previous findings in both real space [11] and virtual space [20, 30, 21]. Of most interest, however, was the robust effect of goal scenario on both distance and time estimation, with more marked effects for time estimation. For both of these measures, both the low and high urgency scenarios were associated with larger distance and time estimates than the no scenario control condition. Moreover, there was also a reliable correlation between self-reported level of urgency during the task and time estimates (but not distance estimates). The second experiment showed that the effect of scenario on memory for time and distance cannot be explained by the influence of physiological arousal, induced with the task instructions for the scenarios, as Experiment 2 revealed no physiological changes across the scenarios, nor were there any differences in distance estimates of the reference path.

The effect of goals on time estimation was not found in the Johnson *et al.* [16] study in real space, and the effects of goals on distance estimation were weaker in virtual space, which may indicate that the lack of vestibular and proprioceptive information in VR results in greater attentional resources to attend to time compared to distance. This would however require further investigation.

Although Experiment 2 rules out the view that scenario affects memory for time and distance as a function of changes in arousal or mood state, it remains to be established exactly how goals affect perception of time and distance. One possibility is that the amount of attention participants allocate to their environment when performing the goals is mediated by scenario, consistent with the ‘attentional shift’ hypothesis [25]. This could possibly be tested further by protocol analysis, allowing participants to ‘think aloud’ when they move through their environment.

In summary, the data reported above suggest that goals do indeed influence memory for time and distance in virtual environments. These results both suggest that the reasons why people move round space are fundamental to how that space is experienced and recalled, and that VR provides an excellence platform in which to test spatial cognition.

References

1. Downs, R.M., Stea, D.: Maps in minds. In: Reflections on cognitive mapping. Harper and Row, New York (1977)
2. Tolman, E.C.: Cognitive maps in rats and men. Psychological Review 55, 189–208 (1948)
3. Richardson, A.R., Montello, D.R., Hegarty, M.: Spatial knowledge acquisition from maps, and from navigation in real and virtual environments. Memory and Cognition 27, 741–750 (1999)
4. Rinck, M., Denis, M.: The metrics of spatial distance traversed during mental imagery. Journal of Experimental Psychology: Learning, Memory, and Cognition 30, 1211–1218 (2004)
5. Avraamides, M.: Spatial updating of environments described in texts. Cognitive Psychology (47), 402–431 (2003)
6. Avraamides, M., Loomis, J.M., Klatzky, R.L., Golledge, R.G.: Functional equivalence of spatial representations derived from vision and language: Evidence from allocentric judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition (30), 801–814 (2004)
7. Franklin, N., Tversky, B.: Searching imagined environments. Journal of Experimental Psychology: General (119), 63–76 (1990)
8. Mani, K., Johnson-Laird, P.N.: The mental representation of spatial descriptions. Memory & Cognition (10), 181–187 (1982)
9. Thorndyke, P.W.: Distance estimation from cognitive maps. Cognitive Psychology 13, 526–550 (1981)
10. Briggs, R.: Urban Cognitive Distance. In: Downs, R.M., Stea, D. (eds.) Image & Environment; Cognitive Mapping and Spatial Behavior, 2005 Ed. Aldine Transaction, New Jersey, p. 368 (1973)
11. Sadalla, E.K., Magel, S.G.: The perception of traversed distance. Environment and Behavior 12, 65–79 (1980)
12. Golledge, R.G.: Environmental cognition. In: Stokols, D., Altman, I. (eds.) Handbook of Environmental Psychology. John Wiley & Son, Chichester (1987)
13. Decety, J., Jeannerod, M., Prablanc, C.: The timing of mentally represented actions. Behavioural Brain Research (34), 35–42 (1989)
14. Taylor, H.A., Naylor, S.J.: Goal-directed effects on processing a spatial environment. Indications from memory and language. Memory & Cognition 27, 309–319 (1999)
15. Rosanno, M.J., Reardon, W.P.: Goal Specificity and the Acquisition of Survey Knowledge. Environment and Behavior 31(3), 395–412 (1999)
16. Johnson, A., Coventry, K.R., Thompson, E.M.: The Effect of Goals on Memory for Human Mazes in Real and Virtual Space. In: Taatgen, N.A., van Rijn, H. (eds.) Proceedings of the 31st Annual Conference of the Cognitive Science Society, July 2009, pp. 1752–1757. Cognitive Science Society, Austin (2009)

17. Montello, D.R.: The Perception and Cognition of Environmental Distance: Direct Sources of Information. In: Hirtle, S.C., Frank, A.U. (eds.) COSIT 1997. LNCS, vol. 1329, pp. 297–311. Springer, Heidelberg (1997)
18. Banton, T., Stefanicci, J., Durgin, F., Fass, A., Proffitt, D.: The perception of walking speed in a virtual environment. Presence-Teleoperators and Virtual Environments 14(4), 394–406 (2005)
19. Campos, J., Nusseck, H.G., Wallraven, C., Mohler, B.J., Bülthoff, H.H.: Visualization and (Mis)Perceptions in Virtual Reality. In: Möller, R. (ed.) Tagungsband 10. Workshop Sichtsysteme, vol. (11), pp. 10–14. Shaker, Aachen (2007)
20. Jansen-Osmann, P., Berendt, B.: Investigating distance knowledge using virtual environments. Environment & Behavior (34), 178–193 (2002)
21. Ruddle, R.A., Payne, S.J., Jones, D.M.: Navigating buildings in desk-top virtual environments: Experimental investigations using extended navigational experience. Journal of Experimental Psychology: Applied 3, 143–159 (1997)
22. Zillmann, D.: Excitation transfer in communication-mediated aggressive behavior. Journal of Experimental Social Psychology 7, 419–434 (1971)
23. Adolphs, R., Damasio, A.: The interaction of affect and cognition:A neurobiological perspective. In: Forgas, J.P. (ed.) Handbook of affect and social cognition, pp. 27–49. Erlbaum, Mahwah, NJ, USA (2001)
24. Phelps, E., Ling, S., Carrasco, M.: Emotion facilitates perception and potentiates the perceptual benefits of attention. Psychological Science 17, 292–299 (2006)
25. Bugmann, D., Coventry, K.R.: Action and the Representation of Distance in Cognitive Maps Acquired Through Imagined Traversal: The Development of a New Methodology. Memory & Cognition 36(3), 518–533 (2008)
26. Massaad, F., Lejeune, T.M., Detrembleur, C.: The up and down bobbing of human walking: a compromise between muscle work and efficiency. Journal of Physiology 58(22), 789–799 (2007)
27. Phelps, E.A., Ling, S., Carrasco, M.: Emotion facilitates perception and potentiates the perceptual benefits of attention. Psychological Science (17), 292–299 (2006)
28. Stefanucci, J.K., Storbeck, J.: Don't Look Down: Emotional Arousal Elevates Height Perception. Journal of Experimental Psychology; General 138(1), 131–145 (2009)
29. McKinnon, A., Jorm, A.F., Christensen, H., Korten, A.E., Jacombe, P.A., Rodgers, B.: A short form of the Positive and Negative Affect Schedule: an evaluation of factorial validity and invariance across demographic variables in a community sample. Personality and Individual differences 27, 405–416 (1999)
30. Jansen-Osmann, P.: Using desktop virtual environments to investigate the role of landmarks. Computers in Human Behavior (18), 427–436 (2002)

The Spatial and Temporal Underpinnings of Social Distance

Justin L. Matthews and Teenie Matlock

Cognitive and Information Sciences
University of California
5200 North Lake Road
Merced, California, USA
{jmatthews,tmatlock}@ucmerced.edu
http://cogsci.ucmerced.edu

Abstract. To what extent do people anchor thoughts about social relationships in terms of space and time? Three studies used drawing and estimation tasks to further explore the conceptual structure of "social" distance. In the three studies, participants read short narratives, drew what they imagined happening during the narrative, then estimated both time and distance. In general, results suggest that the conceptual structure of social relationships is linked to thought about space in terms of path drawing and temporal estimation, but not absolute distance estimation. Results are discussed in terms of mental simulation and inter-character interaction.

Keywords: spatial reasoning, distance estimation, temporal judgments, drawing, mental simulation.

1 Introduction

Everyday language is replete with expressions that describe relationships in terms of physical space. This is evident in the domain of friendship. In talking about friends, people readily use statements such as *We have grown close, They stuck together, Bob stood by his side*, or *He leaned on his buddy after he heard the bad news,* to imply familiarity, fondness, confidence, trust, and so on. They also use statements such as *We have drifted apart, They seem distant lately, He turned his back on his friend*, or *Something came between them*, to imply problems in a friendship.

The goal of this research is to explore the connection between physical space and friendship. To what extent do people think about space when they conceptualize friendship? We are especially interested in whether thought about physical space is part of everyday thought about friendship.

It is well known that people describe abstract concepts in terms of their experience with physical space. This reflects an inclination to draw on relatively more basic domains that are grounded in everyday physical or perceptual experience (see [1]; [2]). For example, people think about time in terms of space, which

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 19–31, 2010.

is reflected in the ubiquity of linguistic expressions, such as *June comes before July, The first week of school has just passed*, and *We are approaching a holiday weekend*. In these cases, a source domain of physical space maps structure on to a target domain of time (see [3]; [4]; [5]; [6]; [7]; [8]). People also think about numbers and mathematics in terms of space, as is evidenced by the number line and language such as *The number is higher than eight*, or *Six hundred comes after 599*[9]. They also conceptualize the internet in terms of space, as is seen in the use of expressions such as *Go to my website, I was at your website*, and *We came to a website about bobcats*[10].

The idea that people think of relationships in terms of physical space is intuitively appealing, especially given the ubiquity of linguistic expressions that refer to friendship in terms of spatial relations. This has been discussed at length in cognitive linguistics, including details about cross-domain mapping (e.g., [2]; [11]; [1]) and relations to other metaphors (e.g., emotion metaphors, see [12]. But it has also been discussed in social psychology, primarily in the realm of "social distance". A few studies in cognitive science have explored connections between physical space and similarity. Casasanto (2008) gathered similarity ratings of various stimuli (abstract nouns, unfamiliar faces, line drawings) under different conditions and found that when stimuli items were placed close to one another, pairs of stimuli were judged as more similar during conceptual judgments and less similar during perceptual judgments. Distance effects have also been found in variations of the Stroop (1935) task. Bar-Anan, Liberman, Trope, and Algom (2007) found the spatial location of words affects categorization time when the words have temporal, social, and/or hypotheticality interpretations, but not for words that lack such semantic properties. Accessing the spatial location of a psychologically distant word such as "others" is faster when the word is presented in the "background" of an image rather than the "foreground" of the same image. The exact opposite accessibility pattern occurs with words that are psychologically proximate, like "we".

Research on spatial mental models has revealed that in some ways imagined space is analogous to physical space[16] [17]; [18]. For instance, when asked to read a narrative about a person moving through a spatial environment, such as a house, people are quicker to access information about objects that are physically proximate to the protagonist than objects that are physically distant[19]. Spatial mental models are known to structure the comprehension of language about imagined environments, including language about motion events both real [20] and fictive [21].

Although little work has addressed the issue of whether social space is directly conceptualized in terms of physical space, attention has been given to the connection between attitudes about nationalities and how they influence relative position in physical space. Burris and Branscombe (2005) recently found that when presented with a variety of cities, one in Canada and one in Mexico, Americans overestimate distance between the cities when one city is in the United States and the other in a foreign location, when compared to two equidistant U.S. cities. This result replicated with Canadians as well, and only held when the estimating party was nationally tied to one of the countries used. This suggests that spatial thinking is involved when participating in on-line distance estimation tasks involving nationality.

In the current work we designed three studies to examine how thought about friendship interacts with thought about physical space. Drawing on our previous work[23]; [24], we created a novel task to explore this relationship across three modes of transportation (walking, driving, and riding) and three measurement variables (drawing, temporal judgment, and distance estimation). Our earlier work results suggested that greater social distance is associated with greater physical distance.

2 General Method

Overview. Here, three drawing and estimation experiments addressed the hypothesis that spatial thinking is related to thought about social relationships, specifically friendship. In these experiments participants first read narratives that described a person either walking, driving, or riding through a park to deliver a package while passing various figures. They then drew the route they would take to accomplish the goal, estimated how much time had passed during the trip, and how far they had traveled (see Figure 1). If thinking about space and social relationships are related, the social relationship described in the narrative should influence the route taken to complete the package delivery task as well as the physical and temporal judgments made about the journey.

Stimuli. The task appeared on a single page in a booklet that consisted of unrelated materials. Participants read a second-person narrative describing a journey through a city park where the reader passes by different people. Half read a passage describing the other people in the park as strangers: "Imagine you need to deliver a package. Along the way, you (walk/drive/ride) through a park and pass by different people. You *do not know* these people. They are *strangers*." The other half of participants read a passage describing the other people in the park as friends: "Imagine you need to deliver a package. Along the way, you (walk/drive/ride) through a park and pass by different people. You *know* these people well. They are *your friends*." Narratives were phrased such that interaction with the other people was not explicitly mentioned. Below the passage, the following instructions were given to participants: "Please draw the route you take through the park using a continuous line". Below the instructions, a simple map (see Figure 1) was given for participants to sketch a path from two points labeled "Start" and "Finish".

2.1 Figures

Maps contained three horizontal rows of trees and/or fencing with a stick figure or vehicle at the end of each tree/fence row. Maps were constructed so a single path from start to finish served as the only solution to the task, and forced participants to pass by the three figures as mentioned in the narrative (see

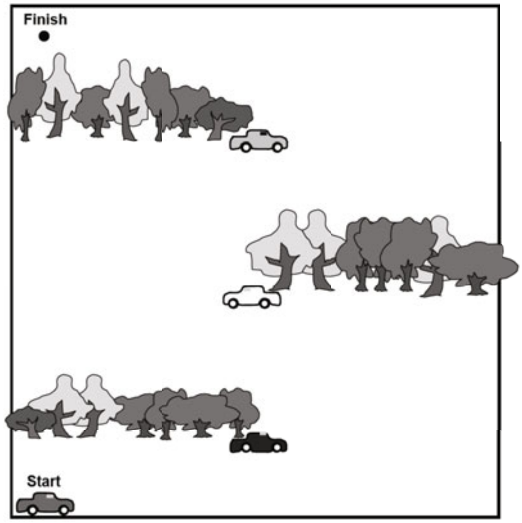


Fig. 1. Example of visual stimuli presented to participants in the *Driving* condition

Figure 1). To complete the task, participants simply drew a continuous line from "Start" to "Finish" depicting their route through the park. After the drawing task, participants provided written estimates of elapsed time: "Using your best guess, how much time (in minutes) did it take you to (walk/drive/ride) through the park?" and distance traveled: "Using your best guess, how far (in feet) do you think you (walked/drove/rode) in the park?" All scenes were similar with only the figures changing across condition (see Figure 2).

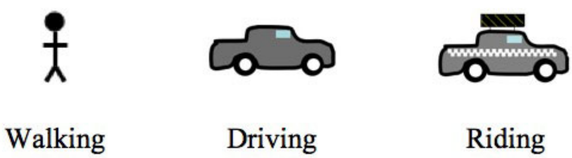


Fig. 2. Figures used in the *Walking, Driving*, and *Riding* conditions

Procedure. Overall, participants were randomly assigned to one of six conditions, differing solely by narrative type: two social relationship conditions (friend or stranger) crossed with three modes of transportation conditions (walking, driving, and riding in a taxi). Participants were instructed to read the narrative then complete the route drawing task and the time and distance estimation tasks.

Data Analysis. The primary dependent variable, distance of the route drawn to the figures in park, was operationalized by measuring (in millimeters) the absolute distance from the drawn route to the most distal end of the arm of each stick figure or the outer edge of the front bumper of the depicted vehicle. Three separate measurements were taken from each participant's drawing (distance from drawn route to the bottom figure, distance from drawn route to the middle figure, and distance from drawn route to the top figure). Drawn routes were also coded for route-figure intersection and route-tree/fence intersection. For analysis, all time estimates were converted to minutes and distance estimates were converted to feet.

3 Experiment 1 – Walking

This experiment examined whether social information influences thought about space and time. Participants were instructed to read a short narrative and imagine walking through a park to deliver a package while plotting their route on a provided map. Friend or stranger figures were positioned along the route. If social information influences thought about space and time, then a difference in route-figure distance should be seen across the social relationship conditions; where routes are drawn closer to friends than to strangers.

Participants. A total of 263 UC Merced undergraduate students (159 women; Age M=18.49, SD=1.09) enrolled in either a Cognitive Science or Psychology course participated for partial course credit.

Results - Drawing. On average, more path-figure intersections were found in the friends condition (M=.22, SD=.74) than in the strangers condition (M=.05, SD=.44), $F(1,259)=5.86$, $p=.02$, $\eta^2=.02$. No reliable differences between the friends condition and the strangers condition were found with regard to walking routes intersecting tree/fence barriers, F<1. No other effects were found, all ps>.05.

The three figure-path measurements were not consistent in distance across position, Wilks' $\lambda=.94$, $F(2,524)=7.95$, $p<.001$, $\eta^2=.06$. Therefore the three positions were independently analyzed. As predicted, participants who read the friends narrative drew their walking routes reliably closer to the figures in the park than did those who read the strangers narrative (friends and strangers respectively) when drawing at the top (M=14.25, SD=12.35; M=22.59, SD=16.95), middle (M=15.63, SD=16.08; M=24.55, SD=18.19), and bottom (M=13.73, SD=12.93; M=21.14, SD=16.03) positions. Wilks' $\lambda=.92$, $F_{top}(1,259)=19.26$, $p<.001$, $\eta^2=.07$ (see Figure 3); $F_{middle}(1,259)=16.90$, $p<.001$, $\eta^2=.06$ (see Figure 4); $F_{bottom}(1,259)=17.66$, $p<.001$, $\eta^2=.06$ (see Figure 5). No other effects (main effect for gender or subsequent higher order interactions) were sanctioned for further examination by the omnibus MANOVA.

Results - Estimation. Data from 10 participants were removed from subsequent analysis for providing time and distance estimates greater than three standard deviations from their respective group means. Participants in the friends condition estimated that it took more time (in minutes) to walk through the park (M=19.43, SD=14.19) than participants in the strangers condition (M=11.56,

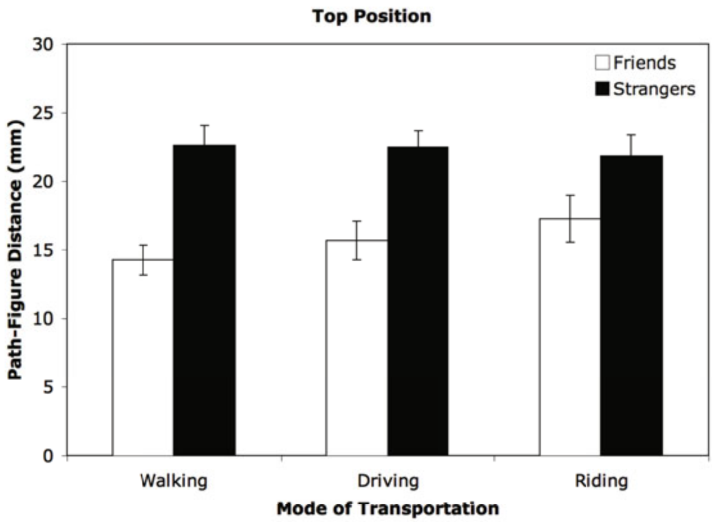


Fig. 3. Top *path-figure distances* (in mm) by *mode of transportation* and *narrative type*

SD=7.01), t(186.68)=5.65, p<.001 (see Figure 6). No reliable differences were found when estimating distance walked between the friends (M=308.35, SD= 459.85) and strangers (M=313.03, SD=534.59) conditions, t(255)=-0.08, p=.94 (see Figure 7).

4 Experiment 2 – Driving

This experiment investigated whether social information influences thought about space and time, with moderate difficulty of actual character interaction taking place. What will happen if the participants imagine riding in cars past other figures in cars? Delivering the package in a car should make it especially difficult to imagine interacting with others in cars. The procedures used were identical to those in Experiment 1, as were the predictions.

Participants. A total of 324 UC Merced undergraduate students (199 women; Age M=20.33, SD=2.72) enrolled in either a Cognitive Science or Psychology course participated for partial course credit.

Results - Drawing. On average, more path-figure intersections were found in the friends condition (M=.52, SD=1.01) than in the strangers condition (M=.23, SD=.72), F(1,320)=7.17, p=.008, η^2=.02. No reliable differences between the friends condition and the strangers condition were noted with regard to driving routes intersecting tree/fence barriers, F<1. No other effects were noted, all ps>.05.

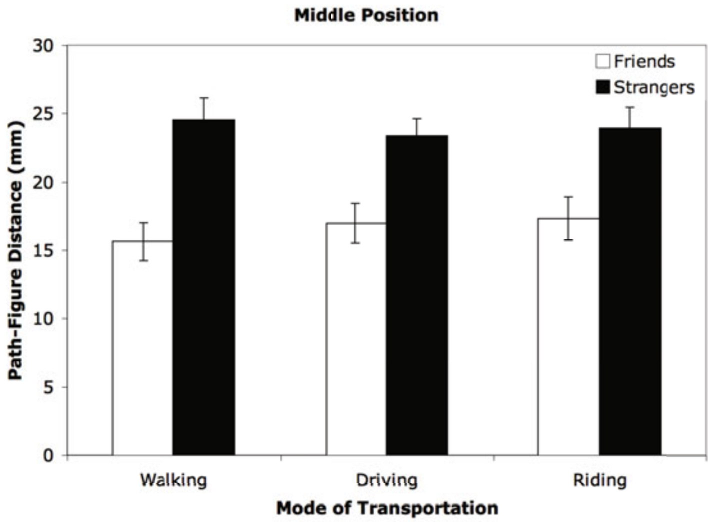


Fig. 4. Middle *path-figure distances* (in mm) by *mode of transportation* and *narrative type*

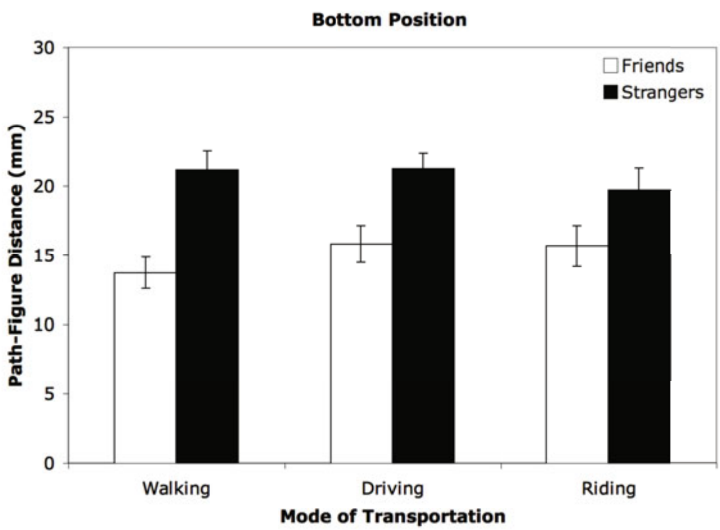


Fig. 5. Bottom *path-figure distances* (in mm) by *mode of transportation* and *narrative type*

The three figure-path measurements were not reliably different from one another with regard to distance across position, Wilks λ=.99, $F(2,322)=2.48$, p=.13, η^2=.01. However methodological consistency was preserved and the three positions were independently analyzed, just as was done in experiment one. As predicted,

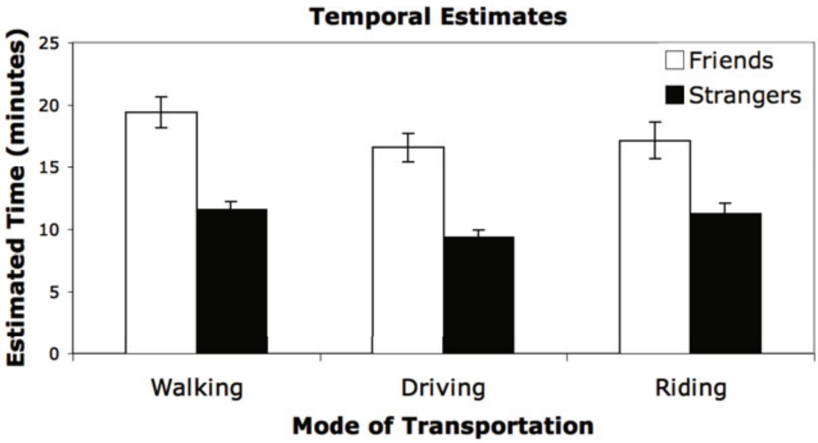


Fig. 6. *Temporal estimations* (in minutes) by *mode of transportation* and *narrative type*

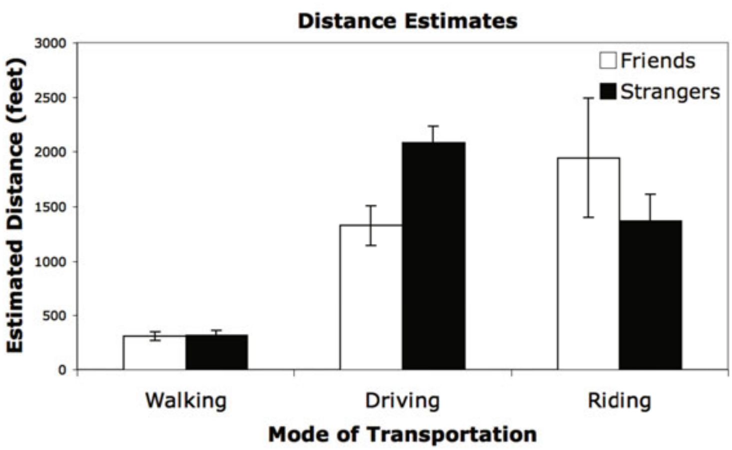


Fig. 7. *Distance estimations* (in feet) by *mode of transportation* and *narrative type*

participants who read the friends narrative drew their driving routes reliably closer to the figures in the park than did those who read the strangers narrative (friends and strangers respectively) when drawing at the top (M=15.69, SD=16.94; M=22.47, SD=16.52), middle (M=16.97, SD=17.33; M=23.39, SD=16.91), and bottom (M=15.82, SD=15.73; M=21.22, SD=15.46) positions, Wilks λ=.94, $F(3,318)=6.82$, $p<.001$, $F_{top}(1,320)=18.30$, $p<.001$, $\eta^2=.05$ (see Figure 3); $F_{middle}(1,320)=14.87$, $p<.001$, $\eta^2=.04$ (see Figure 4); F_{bottom} (1,320)=12.86, $p<.001$, $\eta^2=.04$ (see Figure 5). Men drew their driving routes reliably closer to the figures in the park than did women (men and women respectively) when drawing at the top (M=18.64, SD=16.75; M=20.01, SD=17.20), middle (M= 18.66, SD=16.57; M=21.74, SD=17.79), and bottom (M=16.34, SD=14.01; M=20.40, SD=16.65) positions, Wilks λ=.97, F(3,318)=3.21, p=.02, η^2=.03, F_{top}(1,320)= 2.63, p=.11, η^2=.01; F_{middle}(1,320)=5.12, p=.02, η^2=.02; F_{bottom} (1,320)=8.22, p=.004, η^2=.03. No higher order interactions were sanctioned for further examination by the omnibus MANOVA.

Results - Estimation. Data from 11 participants were removed from subsequent analysis for providing time and distance estimates greater than three standard deviations from their respective group means. Participants in the friends condition estimated that it took more time (in minutes) to drive through the park (M=16.57, SD=13.70) than participants in the strangers condition (M=9.34, SD=7.32), t(202.02)=5.67, p<.001 (see Figure 6). Participants in the friends condition (M=1323.91, SD=2143.22) estimated their traveled distance to be shorter than those in the strangers condition (M=2081.29, SD=4287.92), t(270.56)=-2.05, p=.04 (see Figure 7).

5 Experiment 3 – Riding

This experiment attempted to address whether social information influences thought about space and time, with extreme difficulty of actual character interaction taking place. Here participants imagined delivering a package by riding in a taxi through a park. Again, the procedures used were identical to those in Experiments 1 and 2, as were the predictions.

Participants. A total of 190 UC Merced undergraduate students (115 women; Age M=19.11, SD=1.67) enrolled in either a Cognitive Science or Psychology course participated for partial course credit.

Results - Drawing. On average, more path-figure intersections were found in the friends condition (M=.56, SD=1.12) than in the strangers condition (M=.16, SD=.62), F(1,186)=5.79, p=.02, η^2=.03. On average, more barrier-path intersections were observed in the friends condition (M=.03, SD=.18) than in the strangers condition (M=.00, SD=.00), F(1,186)=4.83, p=.03, η^2=.03. No other effects (main effect for gender or subsequent higher order interactions) were observed for neither path-figure intersections nor barrier-path intersections, all ps>.05. The three figure-path measurements were not consistent in distance across position, Wilks λ=.95, F(2,188)=5.87, p=.007, p=.003, η^2=.03, therefore the three positions were independently analyzed. Participants who read the friends narrative tended to draw their riding routes closer to the figures in the park than did those who read the strangers narrative (friends and strangers respectively) when drawing at the top (M=17.29, SD=16.67; M=21.83, SD=15.00), middle (M=17.32, SD=15.68; M=23.93, SD=14.83), and bottom (M=15.67, SD=14.40; M=19.69, SD=15.35) positions. Wilks' λ=.96, p=.048, F_{top}(1,186)= 2.71, p=.10, η^2=.01 (see Figure 3); F_{middle}(1,186)=7.78, p=.006, η^2=.04 (see Figure 4); F_{bottom}(1,186)=3.05, p=.08, η^2=.02 (see Figure 5). No other effects (main effect for gender or subsequent higher order interactions) were sanctioned for further examination by the omnibus MANOVA.

Results - Estimation. Data from 10 participants were removed from subsequent analysis for providing time and distance estimates greater than three standard deviations from their respective group means. Participants in the friends condition estimated that it took more time (in minutes) to ride through the park (M=17.13, SD=14.35) than participants in the strangers condition (M=11.21, SD=8.11), $t(147.60)$=3.471, p=.001 (see Figure 6). No reliable differences were found when estimating distance rode between the friends (M=1944.21, SD= 5246.91) and strangers (M=1361.65, SD=2361.44) conditions, $t(126.41)$=0.97, p=.33 (see Figure 7).

6 Discussion

Three studies investigated whether an underlying spatial framework influences thinking about social groups. Questions addressed whether or not people think about and express social relationships (social distance) in terms of actual space (physical distance). In all experiments, social relationship primes influenced the way people conceptualized the passage of time and how people drew paths through physical space to complete a task.

In all three experiments, participants drew lines on maps to show where they would travel when delivering a package and estimated how long it would take. In Experiment 1, when participants imagined walking, they drew their lines closer to figures in the park map and estimated that the trip took longer when they believed the figures were friends (versus strangers). Experiment 2 showed the same results. In that case, participants depicted driving routes, specifically, routes past cars that contained friends or strangers. The same effect was also obtained in Experiment 3, in which participants imagined riding in a taxi through the park. Thus, whether the person walked, drove, or rode, social information appeared to influence how people implicitly conceptualized space relative to another person. Friendship consistently resulted in closer distance and longer time.

Based on the data collected, social distance does appear to involve thought about real distance. However, simply mapping thought about physical space onto thought about social space does not represent the entire picture. Both path drawing and temporal judgments are influenced by social relationship primes, at least in the studies presented here. Classic social distance theory posits that social distance and feelings of favorability are inversely related; as the social distance between two groups increases, feelings of mutual favorability between the groups decrease ([25]; [26]; [27]). This relationship was elicited in the three experiments reported here. In general, while drawing an imagined route through an open space populated with figures or vehicles, participants under the impression that the figures or vehicles were their friends tended to draw paths closer to those pictured figures, so close in fact that many routes drawn in the friend condition physically intersected the figures/vehicles in space. Route-figure intersections were 2 to 3 times more prevalent in the friend conditions than in the stranger conditions, even when narratives used in the task never explicitly mentioned figure interaction (see Figures 8 and 9). Are readers simulating interaction with their friends even when interaction is not mentioned? Data from Experiment 1 would support such a conclusion, evidenced by time and distance estimates made by participants after completing the task. When asked to estimate how long it took them to walk through the park, and how far they walked during their travel, estimates regarding distance did not reliably differ by narrative condition, however estimates of time did vary across narrative condition. Participants under the impression they were passing their friends judged the time taken to walk though the park as longer than those in the stranger condition. However, unlike Experiment 1, where the physical qualities of the environment are conducive to character interaction, Experiments 2 and 3 were designed to make potential interaction increasingly more difficult to imagine, yet participants still drew their paths closer to friends figures/vehicles and further from strangers figures/vehicles.

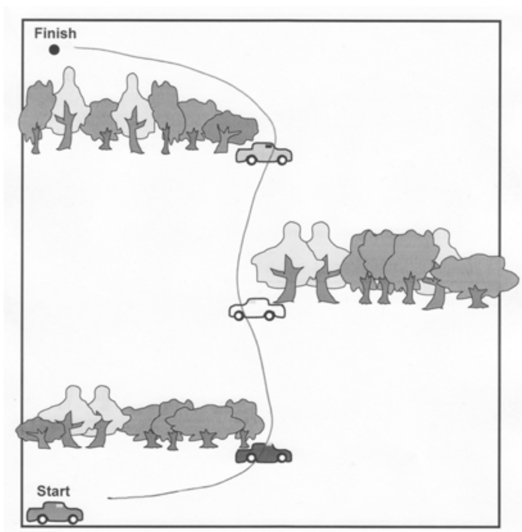


Fig. 8. Participant drawing from the *friend/driving* condition

Future research will address how reasoning about space in the understanding of relationships unfolds over time, including collecting information regarding speed of travel. When people pass by friends, will they slow down, and if so, how much and at what point? Here we only examined situations where participants imagined moving. It will also be informative to explore situations in which others move past the participant (where the stationary figure is the participant). Future explorations in the realm of social distance will also include a variety of manipulations based on social categories ubiquitously found in everyday language, like race, sexual orientation, gender, and ethnicity.

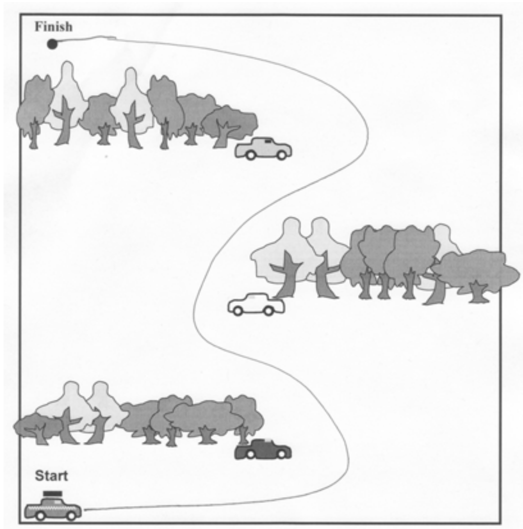


Fig. 9. Participant drawing from the *stranger/driving* condition

In sum, this research examined the interplay between social distance and physical distance, two concepts that have been studied largely by independent groups of researchers, social psychologists and cognitive scientists. The assumption that "distance" in social distance is analogical or metaphoric should be updated to include spatial thinking as a factor that influences the perception of social groups.

Acknowledgments. We would like to express our appreciation to Yarrow Dunham, Bob Levine, Dan Montello, Michael Spivey, Michael Romano, and Michelle Greenwood for insight regarding the findings of this project. We would also like to thank Samantha Bryant, Danielle Birkley, Dawn Felton, Kimberly Williams, Marina Banuelos and Daniel Titcher for their assistance with coding and data entry.

References

1. Gibbs, R.: The poetics of the mind: Figurative thought, language, and understanding. Cambridge University Press, New York (1994)
2. Lakoff, G., Johnson, M.: Metaphors we live by. Chicago University Press, Chicago (1980)
3. Boroditsky, L.: Metaphoric structuring: Understanding time through spatial metaphors. Cognition 75, 1–28 (2000)
4. Clark, E.V., Clark, H.H.: When nouns surface as verbs. Language 55, 767–811 (1979)
5. Gentner, D.: Spatial metaphors in temporal reasoning. In: Gatis, M. (ed.) Spatial schemas in abstract thought. The MIT Press, Cambridge (2001)
6. Gentner, D., Imai, M., Boroditsky, L.: As time goes by: Evidence for two systems in processing space time metaphors. Language and Cognitive Processes 17, 537–565 (2002)

7. McGlone, M.S., Harding, J.L.: Back (or forward?) to the future: The role of perspective in temporal language comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition 24, 1211–1223 (1998)
8. Nunez, R.E., Motz, B.A., Teuscher, U.: Time after time: The psychological reality of the ego- and time reference- point distinction in metaphorical construals of time. Metaphor and Symbol 21, 133–146 (2006)
9. Lakoff, G., Nunez, R.: Where mathematics comes from: How the embodied mind brings mathematics into being. Basic Books, New York (2000)
10. Maglio, P.P., Matlock, T.: The conceptual structure of information space. In: Munro, A., Benyon, D., Hook, K. (eds.) Social navigation of information space, pp. 155–173. Springer, Heidelberg (1999)
11. Lakoff, G., Johnson, M.: Philosophy in the flesh: The embodied mind and its challenge to western thought. Basic Books, New York (1999)
12. Kovecses, Z.: Metaphor and emotion. In: Gibbs Jr., R.G. (ed.) The Cambridge handbook of metaphor and thought, pp. 380–396. Cambridge University Press, New York (2008)
13. Casasanto, D.: Similarity and proximity: When does close in space mean close in mind?. Memory and Cognition 36, 1047–1056 (2008)
14. Stroop: Studies of interference in serial-verbal reaction. Journal of Experimental Psychology 18, 643–662 (1935)
15. Bar-Anan, Y., Liberman, N., Trope, Y., Algom, D.: Automatic processing of psychological distance: Evidence from a Stroop task. Journal of Experimental Psychology 136, 610–622 (2007)
16. Rinck, M., Hahnel, A., Bower, G.H., Glowalla, U.: The metrics of spatial situation models. Journal of Experimental Psychology: Learning, Memory, and Cognition 23, 622–637 (1997)
17. Rinck, M., Denis, M.: The metrics of spatial distance traversed during mental imagery. Journal of Experimental Psychology: Learning, Memory, and Cognition 30, 1211–1218 (2004)
18. Rinck, M., Bower, G.H., Wolf, K.: Distance effects in surface structures and situation models. Scientific Studies of Reading 2, 221–246 (1998)
19. Morrow, D.G., Greenspan, S.L., Bower, G.H.: Accessibility and situation models in narrative comprehension. Journal of Memory and Language 26, 165–187 (1987)
20. Morrow, D.G., Clark, H.H.: Interpreting words in spatial descriptions. Language and Cognitive Processes 3, 275–291 (1988)
21. Matlock, T.: Fictive motion as cognitive simulation. Memory and Cognition 32, 1389–1400 (2004)
22. Burris, C.T., Branscombe, N.R.: Distorted distance estimation induced by a self-relevant national boundary. Journal of Experimental Social Psychology 41, 305–312 (2005)
23. Matthews, J.L., Matlock, T.: How spatial is social distance? Poster presented at the 29th Annual Cognitive Science Society, Nashville, TN (2007)
24. Matthews, J.L., Matlock, T.: Effects of social information on distance estimation. Poster presented at the 30th Annual Cognitive Science Society, Washington, D.C (2008)
25. Bogardus, E.S.: A social distance scale. Sociology and Social Research 17, 265–271 (1933)
26. Hoxter, A.L., Lester, D.: Social distance evaluations in white and African-American students. Perceptual and Motor Skills 80, 478 (1995)
27. Lee, S.Q., Templer, D.I., Mar, J., Canfield, M.: Social distance and trait attribution among four Southeast Asian ethnic groups in the United States. Psychological Reports 91, 326–330 (2002)

The Role of Slope in Human Reorientation

Daniele Nardi, Nora S. Newcombe, and Thomas F. Shipley

Department of Psychology
Temple University
1701 N. 13th Street
Philadelphia, PA 19122, USA
daniele@temple.edu

Abstract. Studies of spatial representation generally focus on flat environments and visual stimuli. However, the world is not flat, and slopes are part of many natural environments. In a series of four experiments, we examined whether humans can use a slope as a source of allocentric, directional information for reorientation. A target was hidden in a corner of a square, featureless enclosure tilted at a 5° angle. Finding it required using the vestibular, kinesthetic and visual cues associated with the slope gradient. Participants succeeded in the task; however, a large sex difference emerged. Men showed a greater ability in using slope and a greater preference for relying on slope as a searching strategy. The female disadvantage was not due to wearing heeled shoes, but was probably related to a greater difficulty in extracting the vertical axis of the slope.

Keywords: spatial abilities, reorientation, vertical dimension, slope or geographical slant, sex differences.

1 Introduction

Re-gaining a sense of orientation once we have lost track of where we are – a process called *reorientation* – is an essential ability for any mobile animal. It is based on the recognition of one or more key elements of the environment (e.g., a celestial cue, a distinctive building, a sound) that provide a sense of "where we are". In the past 25 years, because of the claim that reorientation is selectively attuned to the geometric shape of the environment (the "geometric module" hypothesis; [1]), most research attention has been focused on visual cues; specifically, on whether human and non-human animals can reorient using a geometric layout of walls, and can integrate this information with other visual feature cues (e.g., landmarks).

One spatial cue that has been neglected is the three-dimensional topography of the land. Most of the research on spatial cognition is carried out on completely horizontal surfaces; however, the world is not flat, and we live and move on surfaces of varying elevation – both in man-made and in natural settings. Therefore, it is crucial to examine how space is represented in vertically extended environment, such as dunes, hills and mountains. Vertical topography is a distinctively salient spatial cue for three main reasons. First, this terrain feature is very stable in time and space, whereas other visual features of the environment may change rapidly in the course of a day (different

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 32–40, 2010.

lighting between day and night) or of a season (leafs fall from trees, snow covers vegetation, rivers dry out in summer); because of this permanent nature, vertical topography is a reliable source of spatial information. Second, navigable surfaces extended in the vertical dimension are salient because movements with a vertical component are generally more effortful compared to movements on a horizontal, flat plane; the energy demand associated with counteracting the force of gravity renders the vertical dimension of space cognitively salient. Third, movements with a vertical component provide a suite of multimodal sensory activations, which differ from locomotion on a horizontal surface, as a consequence of kinesthetic, proprioceptive, vestibular and visual stimuli; this multimodal redundancy renders vertical topography uniquely salient.

The simplest surface extended in the vertical dimension is a slope (also called geographical slant; [2]). This provides an allocentric, directional source of information that can be used for reorientation and goal location. In fact, a navigator walking on a tilted surface can reference compass-like bearings extracted from the slope gradient, the two major ones being the vertical axis (uphill/downhill: the direction of steepest descent) and the orthogonal axis of the slope (left/right: direction of no descent; [3]). A slope can be perceived by the different effort when moving uphill, downhill or sideways (kinesthetic information), by the angles of the joints (proprioceptive information), by the sense of balance (vestibular information), and by visual cues (e.g., the angle of incidence between a tree and a slope is acute on the uphill side and obtuse on the downhill side); we refer to these cues collectively as “slope cues” or “slope information”.

Most of the research on slope as a means for reorientation comes from studies on pigeons. It has been shown that pigeons, just like rats [4], can use a slope to locate a goal in an otherwise featureless environment [5]. Lesion studies suggest that this representation is hippocampal-independent [6], similarly to representations based on visual features [7], whereas a geometry-based representation is hippocampal-dependent [8]. When slope and geometry are set in conflict, pigeons preferentially choose the corner that is geometrically incorrect, but that has a correct position relative to the slope [5]. Importantly, slope seems to guide behavior even when it is rendered less informative than geometry [9], suggesting that slope’s salience is so high that it can compensate for a lower predictive value. These findings argue against a primacy of geometric information [1]; furthermore, because the preferred corner is visually incorrect, they do not support a view-based matching strategy for solving reorientation tasks [10].

In humans, there is evidence that the presence of a slope improves navigation in a virtual environment [3], and that a terrain slope – together with other cues – can provide directional information for locating a goal in a virtual environment [11]. However, no study to date has examined whether humans can reorient simply by using a tilted surface, and if they can do so in a real environment. The difference between virtual and real environments is critical, as slope is a multimodal cue and thus cannot be fully represented by visual stimuli on a computer monitor. Therefore, in view of the overall lack of research on spatial cues other than visual ones, and given the ecological relevance of slope, the first purpose of the study here summarized was to investigate the simple – but crucial – question of whether humans are able to reorient by slope.

In spatial cognition, a male advantage is often reported in psychometric tests of mental rotation and spatial perception [12]. Furthermore, a male advantage has been shown in reorientation by geometric cues in a virtual environment [13], and, when

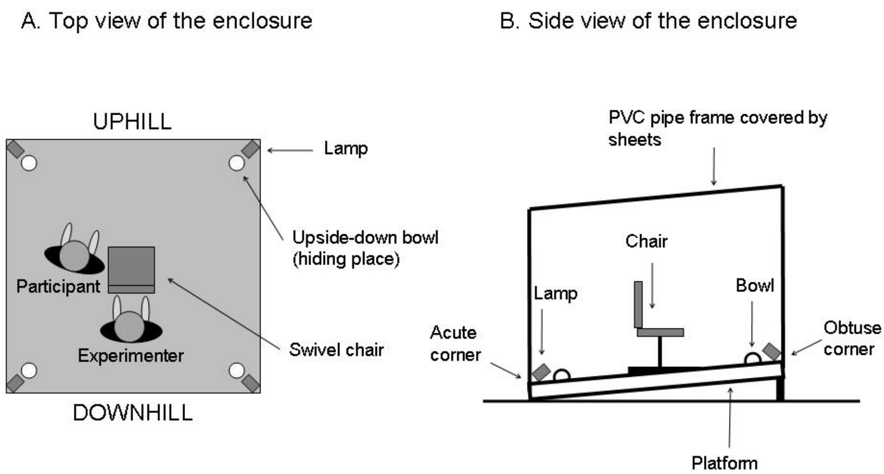


Fig. 1. Schematic representation of the experimental enclosure viewed from above (A) and from the side (B). The size of the enclosure was 244 x 244 cm, and 203 cm high. The enclosure was tilted at an inclination of 5°, but could also be placed horizontally on the ground. In each corner of the enclosure there was a 25-W lamp and a red bowl placed upside-down, which constituted the hiding place for the target (a $1 bill). A swivel chair was placed in the center of the enclosure. When the enclosure was slanted, a wedge was placed under the chair such that the chair's axis of rotation was always parallel to the force of gravity. It is important to note that, when spinning on the swivel chair, the subjects' feet never touched the floor, so no cues were available for keeping track of their position relative to the slope. Each time, a participant entered the enclosure with the blindfold on.

way-finding or navigation abilities are measured, sex differences appear in real-world environments [14] and in virtual environments [15]. Therefore, a secondary goal of the study here summarized was to investigate if sex differences apply to the use of this novel spatial cue – slope.

2 Experiment 1

Twenty male and 20 female Temple University undergraduate students participated. Subjects were shown a target being hidden in one of the corners of a square enclosure (for details see Fig. 1); then they were blindfolded and spun on a swivel chair, to make them lose their sense of orientation. After this, they took off the blindfold, stood up and had to walk to the hidden object. The enclosure was completely featureless, and the square shape did not enable a distinction of the corners based on geometry. However, the enclosure was slanted by a 5° angle – an inclination that pilot studies proved to be easily perceived. Therefore, the goal corner could be identified based on its location relative to the slope gradient (e.g., facing uphill, the goal is on the right; see Fig. 1).

For 4 trials (training phase), the target was always in the same corner (a reference memory paradigm) and participants were given feedback. Both men and women located the target above chance, men: $t(19) = 14.33$, $p < .001$; women $t(19) = 2.41$, $p < .05$. However, men performed significantly better than women (79% vs. 43% correct trials, respectively), $t(38) = 4.43$, $p < .001$, with a difference of 1.4 standard deviations (Fig. 2). After training, two test trials without feedback were carried out to ensure that participants were relying only on slope cues to locate the target. In one trial, the enclosure was placed horizontally on the ground; if participants encoded the goal using slope, now they should be unable to find the target. In a second trial, the enclosure was tilted, just like during training; therefore, participants should be able to use the slope and locate the target. During these test trials (which were in counterbalanced order) both men and women performed at chance when the enclosure was flat (binomial test, males: $p = 1$; females: $p = .80$), and performed above chance when the enclosure was tilted (binomial test, males: $p < .001$; females: $p < .05$), suggesting that, indeed, participants were relying only on slope cues to solve the task.

These results provide the first demonstration that human adults can use terrain slope to reorient and locate a goal in a real-world environment. Using a similar square arrangement of hiding places on a tilted, navigable surface, it has already been shown that rats [4] and pigeons [5] can use the slope to find a goal. Therefore, it was not unexpected that humans could succeed in such a task. However, surprisingly, a large sex difference emerged. Participants could have attempted to use cues other than slope to solve the task (e.g., path integration, details of the enclosure), even though these were ineffective strategies. In this sense, this experiment shows that men have a greater *disposition* to rely on the only cue that consistently predicted the goal – slope – whereas women might attempt to use more other cues. A different question would be: is there a sex difference in the *ability* to use slope? Women could be less disposed, but not less able, to use slope for reorientation.

Fig. 2. Mean percentage of correct trials (± SD) during the 4 training trials in Experiment 1 and in Experiment 3. In Experiment 1 men and women wore casual footwear, whereas in Experiment 3 women wore flat, paper slippers. Men performed significantly more correct trials than women with casual footwear or paper slippers. Women wearing casual footwear and paper slippers performed similarly.

3 Experiment 2

Twenty male and 20 female Temple University undergraduate students participated. The procedure consisted of the same four training trials as in Experiment1 (test trials were not carried out now). However, participants' attention was drawn to the tilt of the floor by showing a ball rolling on the floor of the enclosure. Furthermore, the experimenter suggested that the slope should help in remembering the hiding place. With a clear demonstration that the floor is tilted and an encouragement to rely on it to solve the task, Experiment 2 measured more specifically the ability – rather than the disposition – to use slope information.

The percentage of correct trials increased relative to Experiment 1, $F[1,76] = 15.97$, $MSE = 669.82$, $p < .001$ (Fig. 3). However, men still performed significantly better than women (94% vs. 74% correct trials), $t(38) = 2.44$, $p < .05$, $d = .77$. Comparing performance between Experiment 1 and 2, there was not a significant sex-by-experiment interaction, $F[1,76] = 1.97$, $p = .16$, $\eta^2_p = .03$. Therefore, it seems that a gap between the sexes exists not only in the *disposition* to rely on slope cues (as shown in Experiment 1), but also in the *ability* to use them, because men, when prompted to use slope, performed near ceiling, whereas women still had room for improvement.

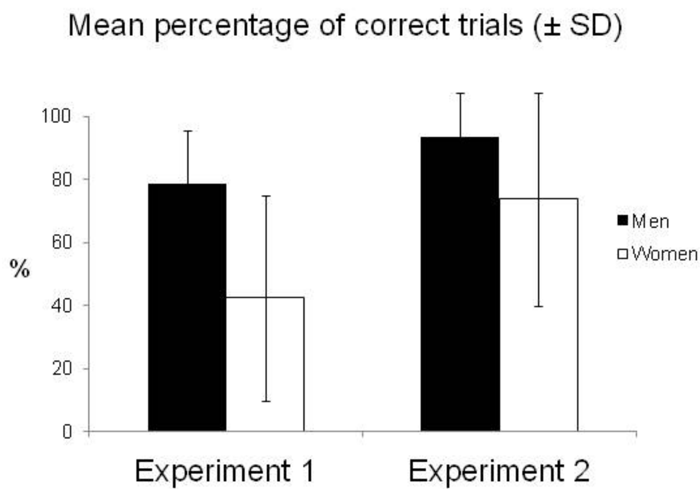


Fig. 3. Mean percentage of correct trials (± SD) during the 4 training trials in Experiment 1 and in Experiment 2. Experiment 1 was an open task, and participants were not given suggestions as to which strategy to use to locate the goal. Conversely, in Experiment 2 attention was drawn to the slope and participants were prompted to rely on it to solve the task. Just like in Experiment 1, in Experiment 2 men performed significantly more correct trials than women. Overall performance in Experiment 2 was significantly higher than in Experiment 1, and the sex-by-experiment interaction was not significant.

4 Experiment 3

What are the factors behind this female disadvantage with slope cues? In Experiment 3 we take into consideration a possibility related to cultural, gender-specific use of footwear. Women's casual footwear is much more likely to present heels, even if moderate, compared to men. Walking on such shoes could introduce a bias in the perception of a sloped floor, rendering kinesthetic and vestibular cues less reliable. This, in turn, could reduce women's sensitivity to slope and impair their ability to use slope for reorientation. In Experiment 3, twenty women were tested in an identical task as in Experiment 1; however, this time they wore disposable paper slippers, which are completely flat. Performance during the four trials was compared with Experiment 1, in which participants wore casual shoes, so that it was possible to test if heels are responsible for the sex gap in reorientation by slope. If women were disadvantaged relative to men because of the heels associated with their casual footwear, their performance in Experiment 3 should be higher than in Experiment 1.

Female performance with paper slippers paralleled that of females with casual shoes in Experiment 1, $t(38) = .12$, $p = .91$ (Fig. 2). The mean percentages of correct trials and the variability were almost identical. It is clear that the footwear worn during the experiment did not have any effect on female performance: women wearing casual shoes or flat, paper slippers performed almost identically and, in both conditions, significantly below men, $t(38) = 4.10$, $p < .001$, $d = 1.30$. Therefore, we can dismiss the possibility that the female disadvantage in Experiment 1 and 2 is due to an adverse effect of female footwear.

5 Experiment 4

In order to use slope for reorientation, one must first be aware of the slope gradient and must be able to extract a direction of reference from it. Experiment 4 examined whether there is a sex difference at these earlier, perceptual stages, as this might be responsible for the female disadvantage using slope for reorientation. Twenty male and 20 female Temple University undergraduate students participated. The same apparatus from previous experiments was used. All participants wore disposable, paper slippers. For three trials participants entered the enclosure and were spun on the swivel chair blindfolded; then they took off the blindfold and were asked a question.

In the first two trials, participants had to judge whether the enclosure was tilted or not. In one trial it was tilted, and in another trial it was placed horizontally on the ground (counterbalanced order). Men and women judged equally well, showing that their awareness of the slope was similar. Then, in the third trial, participants were asked to point as quickly and as accurately as possible to the uphill side of the enclosure. This requires the identification of the vertical axis of the slope, which is considered the most salient direction that can be extracted from the slope, as it is the axis of steepest descent, moving along which most energy is required [5], [16]. Men and women were equally accurate, but men had a significantly shorter reaction time (1.2 vs. 2.3 seconds, respectively), $t(38) = 2.59$, $p < .05$, $d = .82$. Because there was no accuracy/reaction-time trade-off, this suggests that women had more difficulty identifying the uphill direction at the same level of accuracy as men.

Women's difficulty could be perceptual, as they might need more time to process and interpret kinesthetic, vestibular and visual stimuli associated with the slope at a level of confidence. Alternatively, women might perceive the slope as well as men, as suggested by the equivalent accuracies, but might be paying less attention to it, thus requiring more time to attend to slope cues before pointing. Regardless of whether it is a perceptual or attentional difficulty, the longer reaction time in identifying the uphill direction substantiates a female difficulty in dealing with slope cues independent of complex decisions, as there was no goal to remember and no strategy to choose.

6 Conclusions

The present study represents the first demonstration that slope is sufficient for reorientation in humans. This ability is ecologically relevant, as slopes are part of the lay of the land in natural environments. Furthermore, this is the first human study on slope that employs a real environment. When it comes to slope information, the difference between a real and virtual environment is crucial, as this spatial cue provides multimodal stimuli, of which only the visual subset can be presented via computer monitors.

Although both men and women were able to reorient by slope cues, a large male advantage emerged. This can be added to the list of sex differences in spatial cognition shown in mental rotation and spatial perception tasks [12], in reorientation by geometric cues [13], in navigation abilities in real [14] and virtual environments [15]. For most of these differences, there is an advantage in favor of males. However, there is evidence in support of a lack of sex differences in non-visual tasks [2], [17], [18], [19]. One interpretation of these data would suggest that men might have an advantage with visuo-spatial tasks, rather than with spatial tasks *per se*. In this sense, the present study makes a significant contribution to the literature on dimorphic spatial cognition because it shows a robust male advantage in the use of a spatial cue that is not primarily visual, but crucially involves kinesthetic and vestibular stimuli.

Women's disadvantage in using slope is present at the level of perception/attention (Experiment 4), at the level of ability (Experiment 2), and at the level of strategy-preference (Experiment 1). The presence of a sex differences at an earlier stage of information processing (perception and attention) provides a likely explanation for the differences at later stages, when memory of the goal and decision making are involved. Therefore, our data suggest a bottom-up interpretation of the female disadvantage: the difficulty in identifying the vertical axis of the slope gradient can account for the inferior ability in reorienting and locating the goal with respect to the slope; this, in turn, could be the reason why females attempt to use other, ineffective strategies to solve the task.

7 Future Directions

Future studies will be necessary to investigate more deeply the causes of this sex difference. First, it would be interesting to consider if previous experience with slope or, more generally, with directional cues might affect the ability to use slope, in a top-down fashion. Previous studies have suggested that women, compared to men, might be less attuned to directional cues [20]; to what extent can training compensate for this disadvantage?

From Fig. 2 and Fig. 3, it is clear that women had a much larger variability in performance compared to men. This individual difference is linked to the large variety of strategies used: women, more so than men, attempted to rely on other cues to solve the task (e.g., path integration), even though they were ineffective. It would be of extreme interest to find out why some women used slope and some others did not, and if this strategy-preference is correlated to any type of spatial ability (e.g., mental rotation or perspective taking).

Furthermore, one hypothesis that should be contemplated is that the physical build of the participants – and height in particular – may underlie women's disadvantage with slope. The higher the center of gravity of a person, the more likely a small inclination of the floor will require postural adjustments for balance, increasing the awareness of slope. Therefore, men might have outperformed women in the tasks because they are generally taller. Future research will have to take into account physical parameters of participants, and examine if they correlate with performance independently of sex.

Finally, it is possible that women are disadvantaged in the use of slope because of their previous experience with footwear. Experiment 3 showed that footwear worn at the time of the experiment did not alter the perception of the slope. However, women's ability to use slope might be impaired by a history of wearing heels of different height that rendered perceived foot tilt irrelevant. Future research will have to address this issue and investigate if there is a correlation between task performance and footwear habits.

References

1. Cheng, K.: A purely geometric module in the rat's spatial representation. Cognition 23, 149–178 (1986)
2. Proffitt, D.R., Bhalla, M., Gossweiler, M., Midgett, J.: Perciving geographical slant. Psychonomic Bulletin & Review 2, 409–428 (1995)
3. Restat, J.D., Steck, S.D., Mochnatzki, H.F., Mallot, H.A.: Geographical slant facilitates navigation and orientation in virtual environments. Perception 33, 667–687 (2004)
4. Miniaci, M.C., Scotto, P., Bures, J.: Place navigation in rats guided by a vestibular and kinesthetic orienting gradient. Behavioural Neuroscience 113, 1115–1126 (1999)
5. Nardi, D., Bingman, V.P.: Pigeon (Columba livia) encoding of a goal location: The relative importance of shape geometry and slope information. Journal of Comparative Psychology 123, 204–216 (2009)
6. Nardi, D., Bingman, V.P.: Slope-based encoding of a goal location is unaffected by hippocampal lesions in homing pigeons (Columba livia). Behavioral Brain Research 205, 322–326 (2009)
7. Strasser, R., Bingman, V.P.: Goal recognition and hippocampal formation in the homing pigeon (Columba livia). Behav. Neurosci. 111, 1245–1256 (1997)
8. Vargas, J.P., Petruso, E.J., Bingman, V.P.: Hippocampal formation is required for geometric navigation in pigeons. European Journal of Neuroscience 20, 1937–1944 (2004)
9. Nardi, D., Nitsch, K.P., Bingman, V.P.: Slope-Driven Goal Location Behavior in Pigeons. Journal of Experimental Psychology: Animal Behavior Processes (in press)

10. Sturzl, W., Cheung, A., Cheng, K., Zeil, W.: The information content of panoramic images I: The rotational errors and the similarity of views in rectangular experimental arenas. Journal of Experimental Psychology: Animal Behavior Processes 34, 1–14 (2008)
11. Chai, X.J., Jacobs, L.F.: Sex difference in directional cue use in a virtual landscape. Behavioral Neuroscience 123, 276–283 (2009)
12. Voyer, D., Voyer, S., Bryden, M.P.: Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. Psychological Bulletin 117, 250–270 (1995)
13. Kelly, D.M., Bischof, W.F.: Reorienting in images of a three-dimensional environment. Journal of Experimental Psychology: Human Perception and Performance 31, 1391–1403 (2005)
14. Lawton, C.A., Charleston, S.I., Zieles, A.S.: Individual- and gender-related differences in indoor wayfinding. Environment and Behavior 28, 204–219 (1996)
15. Sandstrom, N.J., Kaufman, J., Huettel, S.A.: Males and females use different distal cues in a virtual environment navigation task. Cognitive Brain Research 6, 351–360 (1998)
16. Franklin, N., Tversky, B.: Searching imagined environments. Journal of Experimental Psychology: General 119, 63–76 (1990)
17. Alvis, G.R., Ward, J.R., Dodson, D.L.: Equivalence of male and female performance on a tactuo-spatial maze. Bulletin of the Psychonomic Society 27, 29–30 (1989)
18. Walker, J.T.: Tactual field dependence. Psychonomic Science 26, 311–313 (1972)
19. Berthiaume, F., Robert, M., St-Onge, R., Pelletier, J.: Absence of a gender difference in a haptic version of the water-level task. Bulletin of the Psychonomic Society 31, 57–60 (1993)
20. Jacobs, L.F., Schenk, F.: Unpacking the cognitive map: The parallel map theory of hippocampal function. Psychological Review 110, 285–315 (2003)

Influence of Geometry and Objects on Local Route Choices during Wayfinding

Julia Frankenstein, Simon J. Büchner, Thora Tenbrink, and Christoph Hölscher

SFB/TR 8 Spatial Cognition, Universities of Freiburg / Bremen
{Julia.Frankenstein,buechner,hoelsch}@cognition.uni-freiburg.de,
tenbrink@uni-bremen.de

Abstract. Navigational choices in novel environments are constrained by the wayfinders' expectations on the paths' development beyond the current line of sight. Such expectations may be informed by hallway structures as well as by objects indicating where a hallway may lead to. We study these effects by systematically varying both of these factors in a virtual reality indoor scenario. Results show that earlier findings in both areas gained predominantly in street networks also hold for buildings, and furthermore reveal for the first time how the two factors interact. In particular, attractive objects can cancel the attractive effects of long lines of sight as well as multiple path choices.

Keywords: spatial decision making, indoor navigation, landmarks, virtual environment, geometry.

1 Introduction

Humans are frequently faced with the task of finding new locations within buildings. Although local path choices in such situations are based on incomplete knowledge, they are nevertheless substantially informed by the wayfinder's intuitions and expectations. Often, people "sense" where they might find the main entrance or the auditorium in a public building. Certainly no magic is involved; rather, humans draw on their (often extensive) earlier experience with buildings. As they intuitively understand the concept of a building, they can associate functions with different building parts. Relevant intuitions may be based on a person's experience with specific buildings of various types, or they may represent more generalized background knowledge about common architectural configurations and corresponding visual clues. In spite of the ubiquity of such phenomena and their obvious relevance in everyday life, surprisingly little is known about how indoor design and architectural features actually influence navigation. In this paper, we investigate what kinds of factors constrain local path choices during wayfinding in a previously unknown building. We focus on two fundamental aspects known to be decisive for wayfinding more generally: hallway structure, and object features.

A substantial body of research has addressed the influence of structural characteristics on navigation decisions. Based on a case study in a hospital, Peponis, Zimring, & Choi [1] suggested that patterns of people's movement in a building can be predicted from the features of the building's configuration. More controlled research highlights

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 41–53, 2010.

diverse contributing factors in this regard. For example, the length of line of sight is an important factor for path selection in street networks; people tend to prefer comparatively long views [2][3]. Recent research indicates that this may also be the case in indoor scenarios [4]. For the initial path segment, which is particularly crucial for navigation, people prefer straight lines to curves [5]. Also, paths with many visible future choice options are preferred, as wayfinders aim at maximizing the area they will be able to explore [6]; for eye-tracking evidence see [4]. People furthermore prefer salient axes that may serve as reference directions, supporting their understanding of a building [7]. Apart from pure geometry, factors such as lighting, acoustics, the presence of other people, thermal differences and the like all affect humans' perception of an environment [8] [9][10]. Such effects further contribute to the relative attractiveness of particular paths to choose at a decision point.

In addition to the structural aspects, the cognitive representation of a scene is further affected by the presence of objects, particularly if they are placed at intersections that are relevant for navigation decisions [11]. There is much evidence, mostly accumulated for landmarks outside of buildings, that people use landmarks to orient and locate their own position, to retrace a route, to find the correct direction towards an destination, to describe routes to others, and so on, e.g. [12][13][14]. Objects within buildings may help structuring the environment and shaping navigation strategies [15]. However, the extent to which earlier findings on landmarks may be transferred to indoor scenarios remains unclear, starting from the question of which elements within a building may be perceived as a landmark at all, and how they may differ. In particular, certain objects may be interpreted as cues relevant to the current navigation goal. For example, while a whiteboard may suggest to a wayfinder that this is a place where other people are likely to pass by, a metal door may carry the opposite message, namely to lead to a non-public area. Depending on the current task at hand, such subtle associations and inferences may be more or less supportive for navigation.

Taken together, it can be expected that structural and object features within a building both constrain a wayfinder's choices at local decision points during a navigation task in a previously unknown building. However, only very little research so far has systematically tested each of these effects; for instance, the effect of line of sight has not been addressed with respect to distance size. Moreover, and crucially, these aspects have not yet been directly juxtaposed. Therefore it is unclear to date how they may interact.

In this paper, we address these issues by opposing the role of structural features of a building, such as the length of hallways and number of visible branches, with the role of local, architectural features such as objects and doors that were classified as either attractive with respect to the navigation task, or rather less so. Participants were given the task of finding a publicly accessible central point (the main exit or the auditorium) in a virtual reality building, and confronted with a forced-choice task of local scenes in which structural and architectural features varied systematically. From the considerations and findings as just outlined, we predict that, in the absence of further cues, people will tend to choose longer lines of sight, more future choice options, and paths containing landmarks, in particular attractive ones in relation to the goal. These effects should add to each other if these environmental features are congruent: people should be particularly inclined to choose a path if it offers both an inviting structure, and attractive landmarks. In the case of conflicting cues, adding objects to a configuration should reduce the effects of structure: for example, if a longer line of sight is presented along with non-attractive objects, people should be less inclined to choose this route.

2 Methods

Participants were asked to do a forced choice task between two pictures, each of which showing a view along a hallway in a public building. These hallways differed in geometry (i.e., number of intersections and viewing length) as well as in the presence of different objects.

2.1 Participants

We collected data from 21 participants, ten women and eleven men. They were students of Freiburg University, and they either received course credit or a payment of 8 €/hour for their participation. The total duration of the experiment was approximately 30 min.

2.2 Materials

In order to obtain a sufficient number of wayfinding decisions per participant and to enable within-participant variations, we opted for a forced choice task. Participants were always required to choose between two pictures of hallways that could differ in length of line of sight, number of intersections, and containment of objects.

Task variation. To identify appropriate target destinations as tasks for the main study, a pre-study was conducted as follows. Seven people received a sheet of A4 sized paper with the schematized outline of a rectangular public building. They were asked to indicate where they would look for specific rooms (office of caretaker, auditorium, bathrooms, broom closet, main entrance and exit, director's office, main secretary's office, cafeteria, elevator). From this collection, we identified two locations with a high consensus across participants concerning their position as targets, namely the *main exit* and the *auditorium*. These two targets were chosen for the task variation of the main experiment; thus, participants were asked to find either the main exit (as a prominent way *out* of the building) or the auditorium (as a prominent room *inside* the building). Both of these targets were adequate for our main purpose of providing a clear goal for our participants that could serve as a basis for local navigation decisions, as both suggested publicly accessible and well-known parts of a building.

Stimuli. Virtual reality models of hallway components were built using the open source 3D modeling tool "Blender" (www.blender.org) and a library of free “Google Sketch-up” models of objects. The experiment was programmed in Vizard, a virtual reality animation software. In the experimental program, up to five of the hallway components in different rotations were randomly selected and stuck together, forming hallways with random geometry (see Fig. 1). Then a snapshot of the scene was taken. In each trial, two of these snapshots were presented to our participants on a 17 inch standard computer monitor. Figure 2 shows an example pair of snapshots presented to our participants.

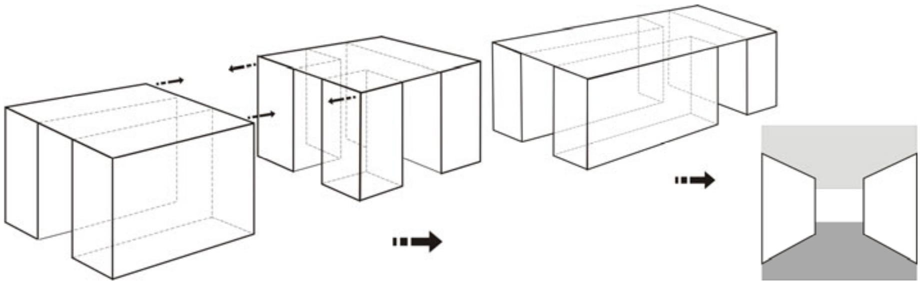
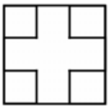
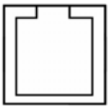
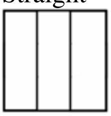
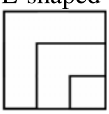
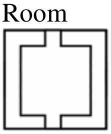
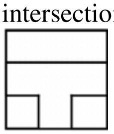


Fig. 1. Different hallway components were randomly selected and stuck together (left side), resulting in hallways differing in shape, viewing length, number of intersections, number of ongoing ways and containing objects. Snapshots of scenes (right side) were used as stimuli.

Fig. 2. One pair of snapshots presented in the experiment

The hallway components differed with respect to shape and containment of objects, but they had the same absolute size, with one exception (a hallway component forming a room, see Table 1). Hallway width was designed to correspond always to four meters in width and three meters in height. The hallway component forming a room was a quadratic open space of 24 x 24 meters and was also three meters high. Illumination and textures were kept constant in all hallway components, and were chosen to match an inconspicuous public building.

Table 1. List of hallway sections used to generate the pictures

Hallway component	Rotation	Objects
Cross inter-section 	0°, 180°	Attractive objects: red sculpture (artwork) only Non-attractive objects: cardboard boxes only without objects
Dead end 	0°	without objects
Straight 	0°	Attractive objects: glass door only, whiteboard only Non-attractive objects: metal door only without objects
L-shaped 	0°, 90°	Attractive objects: glass door only Non-attractive objects: metal door only without objects
Room 	0°	without objects
T-intersection 	0°, 90°, 270°	Attractive objects: white artwork Non-attractive objects: cardboard boxes only without object Combined Objects: glass door + metal door, two rows of chairs, statue + whiteboard, statue + row of chairs

No windows or landmarks except for some closed neutral office doors, a fire extinguisher, an elevator, and the objects used for experimental variation were visible. These objects were selected by three experts considering their plausibility in a public building, and shown in their natural size and color. Furthermore, these objects were pre-classified concerning their attractiveness for the navigation tasks in this experiment. The rationale for this classification was that people looking for a prominent and publicly accessible location (such as the exit and the auditorium) should be attracted by objects that are typical for main hallways in a building (classified as "attractive", e.g., artwork). In contrast, they should be less attracted by objects that are more likely to be found in minor hallways (classified as "non-attractive", e.g., cardboard boxes). We had to find a compromise between the probability to face objects like cardboard boxes in a real public building and the needs of the experiment. So due to our focus on a plausible environment, the number of non-attracting objects is lower than the number of attracting objects, but still high enough to match the needs of the experiment. For detailed combination of hallway components, see Table 1.

For a detailed analysis, we made sure that not all factors between the two pictures varied at the same time. We therefore created five different conditions in which geometry and objects were systematically varied as shown in Table 2. Due to the rotation of the elements, the objects varied in their position.

Table 2. The conditions and their specifications. The columns do not refer to a specific side of presentation (left or right), but present the properties of the two images compared. In condition E the hallway on one side was a mirror image of the hallway on the other side.

Condition	Properties of the hallways presented in separate pictures	
	Picture I	Picture II
A	random geometry, no objects	random geometry, no objects
B	random geometry, one object	random geometry, one object
C	random geometry, one object	random geometry, no objects
D	random geometry, random number of objects	random geometry, random number of objects
E	same geometry, random number of objects	same geometry, random number of objects

2.3 Procedure

Participants were instructed to imagine being in a public building, standing at a T-intersection and having to choose the hallway they would enter to get to the target location (main exit or auditorium). In addition, they were explicitly told to consider the image pairs as independent decisions, rather than attempting to retrace a route or imagine the whole building.

At the beginning of each block, participants watched a short video clip of walking along one hallway without objects and intersections, stopping at a T-intersection. After that, the screen went black for one second, then a pair of snapshots showing two hallways was presented (see Fig. 2). The snapshot itself as well as the location where it was presented (left or right side) were determined randomly. Participants were told that the image on the left side of the screen showed the view towards the left side of the T-intersection, and the image on the right side showed the view to the right. They had to choose the picture of the hallway they would use in order to get to their assigned target location by pressing the arrow key (left or right) of the computer keyboard, and validating their answer by pressing the enter key. The screen turned black again for one second, then the next pair of images was presented. Participants' choices as well as the time they needed for decision and validation were recorded.

Participants performed four blocks of 50 trials, starting with two blocks of the first task (auditorium or main exit) and then performing two blocks of the second task. The order of the two tasks was balanced between participants. Each block contained 50 randomized trials, with ten trials each in the five different conditions (see Table 2).

2.4 Analysis

In some trials the objects presented were not visible to the participants due to the random rotation of the hallway sections. Furthermore, by presenting objects in their natural size we could not avoid occlusion of other objects or changing the line of sight in some trials. Therefore, all trials were recoded to represent the actually observed scene before entering the statistical analysis. To answer the specific questions addressed in this paper, we used some of the different conditions for data calculation. The number of trials resulting from recoding is described for every analysis below.

3 Results

There was a small bias to choose the picture presented on the right-hand side. Over all trials, subjects decided for the right-hand side in 53% of the trials, while in the remaining 47% they chose the hallway presented on the left-hand side. In the following analyses we only report the frequency to choose the right-hand option. The frequency for the left-hand option directly corresponds to it as it was a forced choice task and participants could only either respond “right” OR “left”.

3.1 Geometry

As a first step we analyzed the influence of the hallways’ geometrical features (length of line of sight and number of visible branching paths) on decision behavior. For both data calculations only trials without any visible object in either corridor were analyzed (condition A, $N = 1269$).

In order to test the influence of “length of line of sight” the difference in the length of the lines of sight of both hallways was calculated. The difference was used as a factor in an ANOVA on decision frequency. Length of line of sight had a significant effect on the decision behavior [$F(20,1433) = 10.6$, $p < 0.001$]. Longer hallway sections were chosen over shorter ones (see Fig. 3). Although the increase is not perfectly linear there is a clear pattern.

The same analysis on decision frequency was conducted for the difference in the number of visible branching paths. Their number also had a significant effect on decision frequency [$F(6,1227) = 4.7$, $p < 0.001$]. In general, hallways with a higher number of branching paths were chosen over those with fewer. The analysis shows that both geometrical factors, “length of line of sight” and “number of visible branches”, influence the decision frequency.

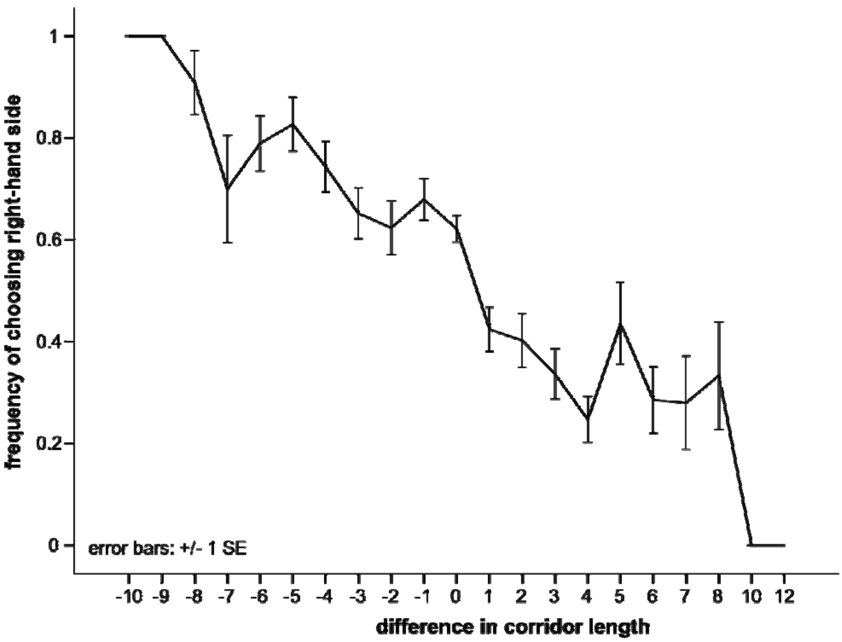


Fig. 3. Frequency of choosing the right-hand option against the length difference of the corridor on the left and the right side. A negative value in the abscissa denotes a longer corridor in the picture presented on the right side, a positive value a longer corridor in the picture presented on the left side.

3.2 Objects

For the analysis of the effect of the presence of objects, only trials in which both geometrical factors were equal on both sides (N = 401) were used. The hallways only differed in the number of objects. The difference in the number of objects on both sides was calculated analogously to the analysis of the geometrical features. An ANOVA revealed a significant main effect [$F(6,394) = 6.2$, $p < 0.001$] resulting from the presence of objects. Participants preferred hallways that contained a higher number of objects to those with fewer objects (see Fig. 4).

3.2.1 Attractive and Non-attractive Objects

For the following analysis the geometry and the number of attractive objects were held constant (N = 389). The two hallways differed only in the number of objects that were pre-classified as potentially non-attractive. An ANOVA with factor “difference in the number of non-attractive objects” was conducted for the decision frequency of the right-hand option. The presence of a non-attractive object had a significant effect on path choice [$F(2,119) = 4.4$, $p < 0.05$]. Participants preferred the hallway with the lower number of non-attractive objects.

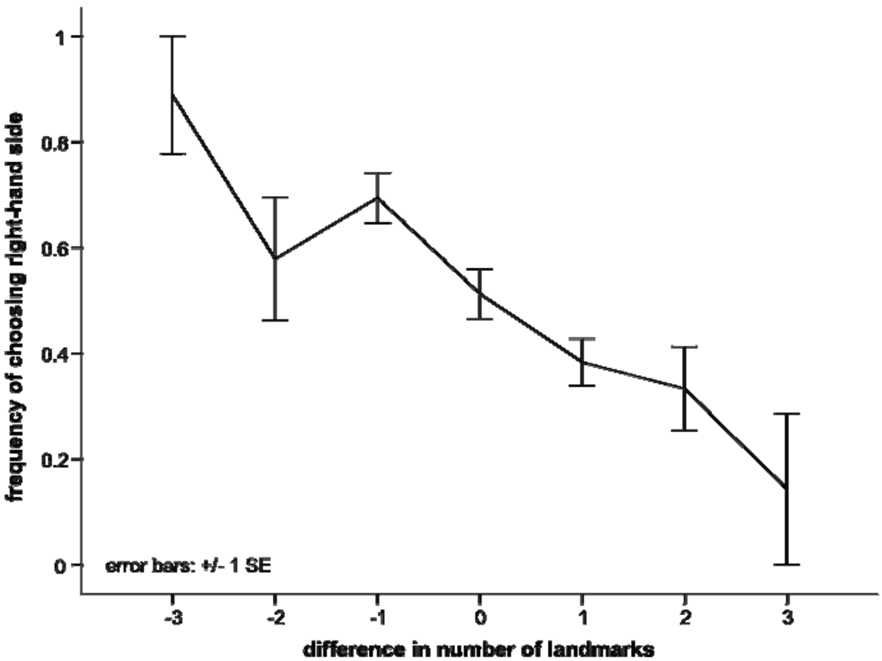


Fig. 4. Probability to choose the right-hand option plotted against the difference in the number of objects in the pictures. A negative value in the abscissa denotes more objects in the picture presented on the right side, a positive value more objects in the picture presented on the left side.

A similar analysis was conducted for the attractive objects. Here the geometry and the number of non-attractive objects were held constant (N = 66). An ANOVA with factor "difference in the number of attractive objects" was conducted for the decision frequency of the right-hand option. The presence of an attractive object had a significant effect on path choice [$F(2,219) = 23.5$, $p < 0.001$]. Participants preferred the hallway with the higher number of attractive objects. The results of these two analyses are displayed in Figure 5.

3.2.2 Direct Comparison of Geometry and Objects

While the previous analyses kept either geometry or objects constant, the following analysis provides initial insights about how the two factors interact. Only trials with a single object on one side and different geometry were analyzed (Condition C, N = 771). This results in four constellations as follows. First, an attractive geometry accompanied by an attractive object is compared to a non-attractive geometry without any object *("attractive, attractive"* constellation). Second, a non-attractive geometry accompanied by a non-attractive object is compared to an attractive geometry without any object (*"non-attractive, non-attractive"* constellation). Third, a non-attractive geometry accompanied by an attractive object is compared to an attractive geometry (*"non-attractive, attractive"* constellation). Fourth, an attractive geometry accompanied by a non-attractive object is compared to a non-attractive geometry (*"attractive,*

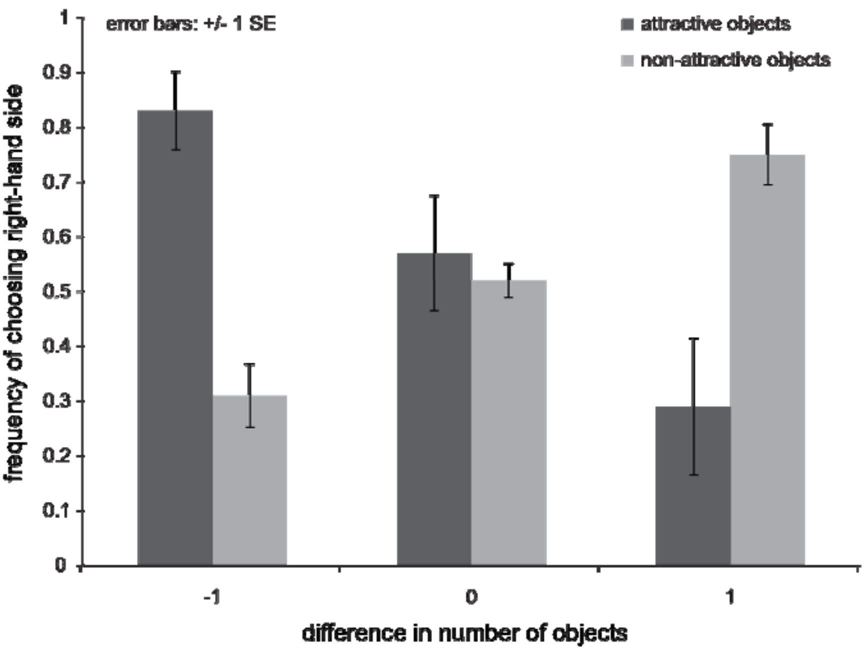


Fig. 5. The frequency of choosing the right hand hallway plotted against the difference in the number of attractive and non-attractive objects in the hallways. The dark bars represent the choice frequency of the trials varying in the number of attractive objects while the number of non-attractive objects was kept constant. The light bars represent the choice frequency of the trials varying in the number of non-attractive objects while the number of attractive objects was kept constant. Negative value in the abscissa denotes more objects were presented on the right hallway, a positive value more objects were presented on the left hallway.

non-attractive" constellation). For this analysis, an attractive geometry was defined by a longer line of sight.

These four different constellations were compared in a single-factor ANOVA. There was a main effect of the constellations [$F(3,767) = 19.8$, $p < 0.001$]. Post-hoc Scheffé tests revealed a significant difference between the "*attractive attractive"* constellation and all other constellations (all $p < 0.001$). In the "*non-attractive non-attractive"* constellation the hallway was chosen less often than in the two incongruent constellations. This difference was significant ($p < 0.05$) concerning the constellation *"attractive non-attractive"*. We found only a trend ($p < 0.1$) for the difference of choice frequency between the "*non-attractive non-attractive"* constellation and the constellation *"non-attractive attractive"*. This is likely due to the low N, because the choice frequency of the two incongruent constellations did not differ (see Fig. 6).

The constellation of an attractive geometry and an attractive object was chosen most frequently. The constellation of a non-attractive geometry and a non-attractive object was avoided most frequently. Most interestingly, the effects of an attractive object and a non-attractive geometry (and vice versa) balanced each other out.

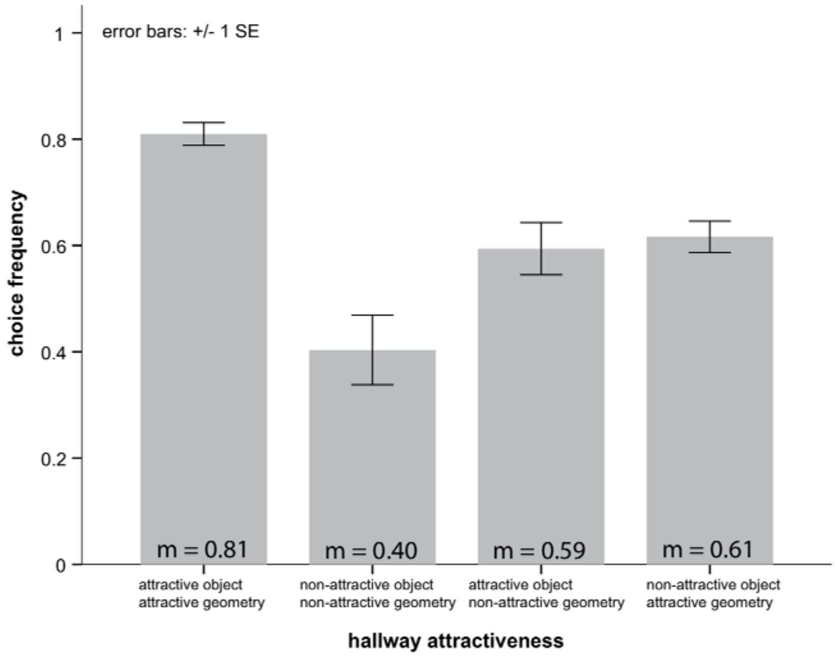


Fig. 6. Choice frequency of hallways with one object and non-/attractive geometry based on line of sight

4 Discussion

The aim of the study was to gain insight into the influence of hallway geometry as well as different types of objects on path choice during wayfinding in a building. Using a forced choice paradigm set in a naturalistic virtual reality environment, we not only systematically varied the presence of particular structures and objects, but also the degree to which they were present, and we addressed the ways in which the individual factors interact. As a result, we were able to support and extend earlier findings about wayfinding (gained predominantly in street networks) for building scenarios. Furthermore, we gained new insights on the graded influence of each addressed factor, and showed that different types of objects influence navigation to different degrees. Most crucially, we evaluate for the first time how geometry and the objects might interact in a situation where both factors become relevant for the wayfinder.

With respect to each individual factor, our results revealed a clear preference for longer lines of sight as well as higher number of visible branching paths. Although the increase is not perfectly linear, the overall pattern does suggest that people prefer to gain as much information about future navigation options as possible. Objects visible on a path generally attracted participants to choose this path; here, as well, the effect was found to be graded. A higher number of objects thus contributes to the salience of a path, which fits well with the given task of navigating towards a prominent, publicly accessible area such as the main exit or the auditorium. However, a more detailed analysis of object types revealed that participants rather avoided hallways with non-attractive objects, but preferred hallways with a higher number of attractive objects. This result corresponds to our previous classification of object types and confirms our prediction that wayfinders attend to the attractiveness of particular objects in relation to the current task.

In order to learn how people react if confronted with various environmental cues at the same time, we addressed the interaction of the two main factors (geometry and objects). Since we could already identify a gradedness effect for each single factor, the combination of different kinds of environmental cues should further support the wayfinders' navigation strategies. Our results clearly support this hypothesis. In the case of congruent cues, namely an attractive structure (longer lines of sight) together with an attractive object (e.g., artwork or a glass door) the corresponding hallway was chosen with a particularly high frequency – the individual effects clearly added up in supporting the wayfinder. Likewise, congruent non-attractive cues led participants to choose the alternative path direction. In the case of incongruent cues in which one of the factors was attractive but the other non-attractive, decisions were near chance level. In other words, a non-attractive object can eliminate the attracting effect of a long line of sight, and vice versa.

Our results regarding geometry as well as the role of attractive and non-attractive objects is clearly in line with the notion of "intelligibility" in the space syntax community. Hillier [2] suggested that people will rely predominantly on major, central hallways for finding their way to novel destinations in a complex building (see also [16]). Buildings with a high intelligibility are those in which the number of local path choices originating from a hallway is strongly correlated with topological centrality ("integration" in space syntax parlance) of that hallway in the building as a whole. This notion provides a theoretical explanation for why people should prefer long lines of sight and a high number of local path choices. In typical public buildings, these local geometric features serve as cues for the relevance of a corridor to efficiently access large parts of the setting. Moreover, the notion is also compatible with the observation that objects associated with central hallways are preferred over objects that are more suggestive of marking a minor, peripheral hallway. In other words, the present study extends the wayfinding heuristics described by the space syntax community for purely geometric properties to semantically rich objects and local architectural features like doors and portals.

This study is an initial stepping stone in a larger research enterprise to understand spatial inference processes in human navigation. When people have only incomplete knowledge about an environment, e.g. after exploring and learning some but not all parts of it, they need to draw inferences from an incomplete cognitive map to venture into unknown areas. While it is an open question to which degree such inferences rely on explicit spatial reasoning, people will certainly make use of any visual and geometric cues that are directly provided by the environment during navigation. In the present study, participants were forced to base their judgements almost entirely on the currently visible information at each decision point, as they had no prior exposure to the building and the experimental procedures discouraged cognitive mapping in favour of local spatial inference. Future studies will systematically investigate how these perceptually driven inferences interact with memory-based wayfinding inferences in active navigation tasks, and with more global wayfinding strategies such as keeping a constant direction of movement [17] or orienting towards a central point [15].

Acknowledgments. The authors thank Philipp Vath and André Guenther for help in 3d modelling and software programming, and Talita Telma for help in running the experiment. Furthermore, the authors want to thank Betty J. Mohler and Jan Peters for their very helpful comments on this paper. Funding by the DFG (SFB-TR 8 Spatial Cognition Bremen / Freiburg) is gratefully acknowledged.

References

1. Peponis, J., Zimring, C., Choi, Y.K.: Finding the building in wayfinding. Environment and Behavior 22(5), 555–590 (1990)
2. Hillier, B.: Space is the machine. Cambridge University Press, Cambridge (1996)
3. Hochmair, H., Frank, A.U.: Influence of estimation errors on wayfinding decisions in unknown street networks - analyzing the least-angle strategy. Spatial Cognition and Computation 2(4), 283–313 (2002)
4. Wiener, J.M., Hölscher, C., Buechner, S.J., Konieczny, L.: How the Geometry of Space controls Visual Attention during Spatial Decision Making. Paper presented at the Annual Meeting of the Cognitive Science Society 2009, Amsterdam, NL (2009)
5. Bailenson, J., Shum, M., Uttal, D.: The initial segment strategy: A heuristic for route selection. Memory and Cognition 28(2), 306–318 (2000)
6. Haq, S., Zimring, C.: Just down the road a piece: The development of topological knowledge of building layouts. Environment and Behavior 35, 132–160 (2003)
7. Werner, S., Long, P.: Cognition meets Le Corbusier - Cognitive principles of architectural design. In: Freksa, C., Brauer, W., Habel, C., Wender, K.F. (eds.) Spatial Cognition III. LNCS (LNAI), vol. 2685, pp. 112–126. Springer, Heidelberg (2003)
8. Canter, D., Stringer, P. (eds.): Environmental Interaction: Psychological Approaches to our Physical Surroundings. International Universities Press, Inc., New York (1975)
9. Hetherington, J., Daniel, T., Brown, T.: Is motion more important than it sounds?: the medium of presentation in environment perception research. Journal of Environmental Psychology 13, 283–291 (1993)
10. Zacharias, J.: Path choice and visual stimuli: Signs of human activity and architecture. Journal of Environmental Psychology 21(4), 341–352 (2001)
11. Janzen, G., Weststeijn, C.G.: Neural representation of object location and route direction: An event-related fMRI study. Brain Research 1165, 116–125 (2007)
12. Presson, C.C., Montello, D.R.: Points of reference in spatial cognition: Stalking the elusive landmark. British Journal of Developmental Psychology 6(4), 378–381 (1988)
13. Lovelace, K., Hegarty, M., Montello, D.R.: Elements of good route directions in familiar and unfamiliar environments. In: Freksa, C., Mark, D.M. (eds.) Spatial information theory. Cognitive and computational foundations of geographic information science, pp. 65–82. Springer, Berlin (1999)
14. Caduff, D., Timpf, S.: On the assessment of landmark salience for human naviga-tion. Cognitive Processing 9, 249–267 (2008)
15. Hölscher, C., Meilinger, T., Vrachliotis, G., Brösamle, M., Knauff, M.: Up the down staircase: Wayfinding strategies in multi-level buildings. Journal of Environmental Psychology 26(4), 284–299 (2006)
16. Dalton, R.C.: Space Syntax and Spatial Cognition. World Architecture 185, 41–47, 107–111 (2005)
17. Dalton, R.C.: The secret is to follow your nose: Route path selection and angularity. Environment and Behavior 35(1), 107–131 (2003)

Testing Landmark Identification Theories in Virtual Environments

Denise Peters[1], Yunhui Wu[2], and Stephan Winter[2]

[1] SFB/TR 8 Spatial Cognition, University of Bremen, Germany
peters@sfbtr8.uni-bremen.de
[2] Department of Geomatics, University of Melbourne, Australia
y.wu21@pgrad.unimelb.edu.au, winter@unimelb.edu.au

Abstract. Landmarks are fundamental elements for people to learn an environment. People use these landmarks to enrich their route descriptions (for example, when anchoring movements at decision points). Several automatic landmark identification and selection theories have been suggested in recent years. This paper evaluates these theories by comparing the landmarks identified by automatic landmark selection with landmark choices and behavioral characteristics of human participants moving in a virtual environment. This comparison of automatic selection and human behavior will improve our understanding of automatic landmark identification theories, and will facilitate a weighting of methods for navigation services to generate more human-like route descriptions.

1 Introduction

Landmarks are fundamental elements for people to learn an environment, i.e., for building up a mental representation that enables people to orient themselves and to navigate in this environment. As a rich body of research in spatial cognitive science and linguistics has shown, people also use landmarks to enrich their route descriptions; for example, for anchoring turning actions at decision points, or as an assurance along longer route segments. Yet current navigation assistance services do not refer to landmarks in their route descriptions because of the lack of reliable methods for identifying suitable landmarks. Furthermore, though several competing automatic landmark identification methods have been suggested in the literature, only one [1] is operational.

All landmark identification methods suggested so far argue for their approach from a cognitive ergonomics point of view (some frame it as information theory) and ground their measures in observed or expected salience. But none of them went through thorough user testing. In this paper we present an experiment that can be used to test and compare any landmark identification method.

This paper addresses two key research questions: (1) do landmark identification methods sufficiently reflect what people refer to in their route descriptions, and (2) does the attention span of people while learning a new route correlate to their choice of landmark in their route descriptions?

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 54–69, 2010.

We are interested in establishing quality measures for landmark selection methods by comparing correlation coefficients of automatic selected landmarks with people's preferences. Accordingly, the hypothesis of this paper is two-fold: (a) a correlation larger than random between landmark selection methods and people's choices of landmarks can be observed, and (b) a correlation larger than random between peoples chosen landmarks and their attention span can be observed.

Our test scenario takes place in a virtual environment. This experimental setup allows us to control the environmental factors easily and precisely record the trajectories during a routing task. The participants navigate along two highlighted routes and write down their own route descriptions for another person to follow. Our analysis consists of three parts: first we will analyze the route descriptions according to the mentioned landmarks, then we will compute the correlation of different landmark identification methods with human-chosen landmarks, and finally we will calculate the correlation between features participants paid attention to and the references in their route descriptions.

As the virtual environment is populated only by empty street space and façades, we expect that participants will use building façades as landmarks, and further, that they will resort to other salient elements, such as street intersections or even geometry where there is a lack of landmarks. We predict a significant degree of unity in their choices where they pass salient landmarks (in terms of the landmark identification methods). We also assume that there will be a correlation between peoples attention in the environment and the chosen landmarks in their route descriptions, which would enable us to suggest new landmark identification models based on tracking attention, and also to evidentially support the selection methods that put some emphasis on visual salience. We further anticipate proof for different influences of landmark identification factors.

2 Landmarks in Route Communication

A landmark is a salient geographic entity that marks a locality and can be used for orienting or navigating in the environment [2,3]. References to landmarks can frequently be found in human route descriptions [4]. Research on spatial cognition and spatial linguistics shows that in route communication people prefer qualitative references over quantitative ones (e.g., [5,6,7]), and accordingly, prefer references to landmarks rather than metric distances (e.g., [8,9,10,11,4]).

Landmarks are mentioned in route descriptions because they are salient in the environment and memorable [7]. Direction givers assume that nominated landmarks are either known or recognizable by listeners. Sorrows & Hirtle [12] state that geographic entities can be memorable for (1) their singularity, or sharp contrast with their surroundings; (2) the prominence of their spatial location; (3) their content, meaning, use or cultural significance; and (4) their prototypicality within a category. Landmarks are frequently categorized as: visual, if they have certain visual characteristics; cognitive, which are semantically meaningful to personal interests or experiences; and structural, which are memorable because of their spatial location [13,14]. Tomko & Winter [15] argue that the prominent references used in route descriptions are brief and can be easily picked up by wayfinders from their cognitive maps of the environment.

2.1 Landmark Identification Theories

A landmark must stand out from its environment in some respect, such as size, color or function, to mark a locality. To enrich navigation services, several theories have been developed for an automatic identification of landmarks from geographic entities representing urban environments. Raubal & Winter [13] present a method to automatically extract landmarks by assessing their salience from the above identified three aspects: visual, semantic and structural attraction. Elias [3], however, suggests a data mining method to select features with outstanding attributes in a neighborhood; the data contains all geometric and thematic aspects of features without consideration of perception by human senses. Again having visual, semantic and structural attraction in mind, Duckham et al. [1] suggest an identification method based on feature categories instead of individual landmarks. In addition to absolute landmarkness, some aspects of landmarks are route-specific, especially advance visibility (e.g., [16,17,18]). This section will study the theories of identifying landmarks focusing on the three aspects of structure, semantics and visibility.

Structural Salience. A landmarks salience can be affected by its position along the route. Michon & Denis [19] study the use of landmarks in route descriptions and conclude that landmarks are used frequently, especially around the starting/arrival point and points where a change in direction is required, i.e., decision points. We use their theory in extracting landmark candidates.

Klippel & Winter [20] formalize the structural salience of point-like objects along a route whereas Claramunt & Winter [21] extend structural salience to elements in street networks – for example, places, paths, barriers and districts – and identify a generic model of structural salience for urban environments.

In our experiment, façades and street intersections are the only elements that can be considered as landmarks, and façades can have structural salience according to their relation to decision points. Other elements, such as barriers or districts, are not provided in this virtual environment.

Semantic Salience. In Winter's [16] survey of the prominence of campus features, the single chapel on campus is more prominent than other landmarks that belong to categories with multiple occurrences on campus, such as bookstores (ranked second, because there were only a few) or halls. Landmarks can stand out by their function as well as their cultural significance, so called semantic salience, independent from visibility or structural salience.

Semantic salience can apply to specific communities – for example, for engineering students the engineering building on campus is more prominent than other buildings, and a suitable landmark for local orientation ("next to our building"). Similarly, personalized route descriptions can even refer to entities with an individual semantic salience ("next to the coffee shop where we met yesterday").

Visual Salience. A landmark may stand out simply for its visible properties, independent of structure or semantics, such as color, size (as being perceivable from street space, i.e., mainly width and height) or form [13]. Among these properties, color is an attribute difficult to formalize. It depends not only on lighting, viewing distance and surrounding colors [22], but also on the texture of the feature.

A method to compute the salience of the color of an image is suggested by Aziz & Mertsching [23], and Nothegger et al. [24] propose an approach for images of façades. Winters [16] measure of advance visibility (how much a façade is visible when approaching a decision point) will also be applied in this paper.

2.2 Landmarks in Virtual Environments

Virtual environments (*VE*) are a valuable tool in environmental psychology. Advantages of VE are the reproducibility of studies, the minimization of interferences, and the accurate recordings of the subjects behavior and performance [25]. A VE is usually an interactive, real-time, 3D graphical rendering of spatial data. Subjects have a first-person egocentric view of the scene and control the interaction by a game-pad or a mouse. They acquire information about an environment sequentially and integrate this information over time into a mental representation of the environment. Many studies show that people are able to learn the spatial layout of a VE and can perform tasks such as route learning successfully.

In fact, a number of studies show that people use landmarks in VEs for their navigation tasks (e.g., [26,27,28]),and further, that the manipulation of landmarks changes peoples navigation behavior.

Additional findings support that it is possible to transfer spatial knowledge from a VE to a real environment, and that a VE can be used to explain a route to a person (e.g., [29,30,31]).

2.3 Interpreting Human Written Route Descriptions

Tversky & Lee [32] and Tenbrink [6] study the relationship between space and language, and find evidence that human written route descriptions reflect the cognitive representations of space while others are concerned with how to generate human-like route descriptions [33,34]. However, there are few studies on how to interpret human written route descriptions, which are normally rich in information and complicated to formalize [35].

In this paper, we focus on the landmarks mentioned in route descriptions. Written references to landmarks are manually identified and annotated without the use of any language technology.

3 Experiment

For our experiment we use a virtual environment that was designed and implemented originally for another purpose [36] and is re-used here with permission of the Hasso Plattner Institut, Potsdam. The virtual environment presents a virtual city with box buildings. Every box building is represented by gluing a photorealistic façade on a cuboid. In this virtual environment, participants see a route highlighted by arrows.

Two routes are available: The short route is 1.66 km long with eight intersections along the route and 38 different façades available to be glued on the box buildings. The other route is 2.8 km long, with 16 intersections and 126 different façades (see Figure 1).

In both cases, the amount of available façade images is not sufficient to texture each building front individually; hence, façade images repeat in a random order. However, the available façade images are of diverse characteristics–they were taken as a representative subset of façades of a Berlin suburb. We use these urban façade images because they offer a high variability of different features, such as shops at the ground level, architectural features and color intensive textures, which can all contribute to landmarkness.

The virtual environment contains no other features of salience beyond the façades: the width of all streets is equal as well as the height of all houses, and no other outstanding objects are placed in the virtual environment.

Fig. 1. This figure shows an overview of the two routes. The short route is part of the long one

Participants will follow these two routes, in random order, and provide a route description for each. We will compare the references to landmarks in the directions we collect to automatically selected landmarks. Also, participants are tracked in the virtual environment with respect to both their movement and their viewing directions (heading). This allows us to analyze how much time people spent looking at each façade. For each route we will determine which façades the participants paid most attention to, then correlate the façades attracting attention to the landmarks referred to in route descriptions. We can thus ascertain whether participants typically refer to those façades they looked at for longer times during active navigation.

24 participants attended the experiment. Participants were students and researchers at the University of Melbourne, 15 males and nine females, aged 23-42. Eight of the participants were of German nationality (which is relevant due to the German context of façade images). All participants were instructed to write down a route description of their traveled route for a friend in daily route communication, with no time limit.

The virtual environment was presented on a large screen projection: Participants sat two meters in front of a 2 by 1.5 meter screen. They navigated actively through the virtual environment with a joystick. Figure 2 shows the starting point for the long route. Participants could navigate with two types of velocity: walking speed or running. The running speed was introduced since the routes cover long distances, which take considerable travel time at walking speed.

In the experiment, all participants navigated twice along each route to explore the routes and collect all information needed for describing them. During the two trials each participant could navigate along the routes at their own speed, and stop and write down their route descriptions at any time. All route descriptions were handwritten in English.

4 Analysis and Results

4.1 Route Descriptions

To analyze the results, we focus on the referred landmarks, their frequency of reference and the manner they are referenced in the directions. In the case of the short route, 23 route descriptions were analyzed (one participant had to be excluded due to technical reasons).

14 out of 23 participants had used landmarks in their descriptions; nine participants had written route descriptions without referring to landmarks, but using chunkings of blocks/intersections and turns. One example of a route description without landmarks is given in Figure 3.

Only three participants had used distance information in their directions. The 14 participants using landmarks had referred to an average of four landmarks for describing the short route, and of these, two were explicit landmarks – meaning a participant named the landmark, such as *Lotto shop* – and two were implicit landmarks – meaning a landmark was described by its properties (e.g., color or architectural structure), such as (*house with balconies* (Figure 4). The most

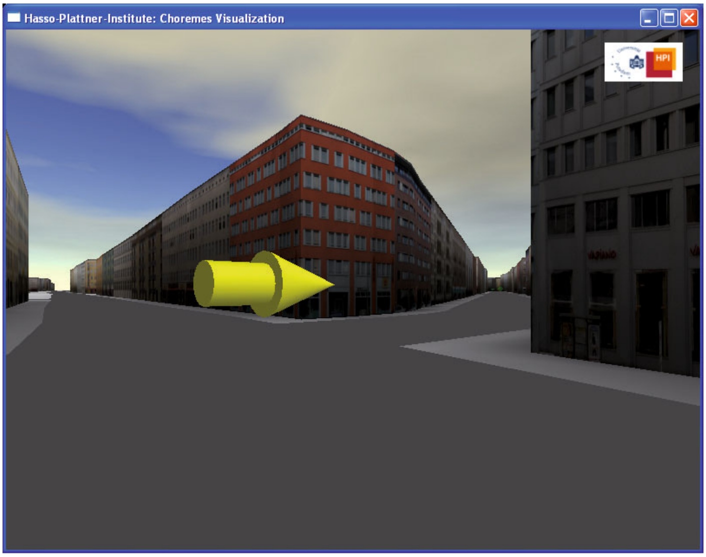


Fig. 2. Starting point of the long route. The arrow shows the direction of the route.

```
Take the street on the right
Go for 3 blocks turn right
First intersection turn right then turn left
Go till the end of the road, turn right
Second left. The end
```

Fig. 3. An example of route descriptions for the short route without using landmarks

prominent explicit landmarks, with ten references each, were the *Lotto shop* and *Ulrich Augenoptik* (Figure 5).

In the route descriptions for the long route 15 participants referred to landmarks in their directions, while nine did not use any. For the participants utilizing landmarks there were, on average, nine landmarks mentioned. An average of two of these landmarks were explicit and seven were implicit. Only four people used distance information. The most prominent landmarks for the long route were a restaurant named *Viapiano* and a *Häagen Dazs* ice cream shop next to this restaurant, both with 11 references (see Figure 5).

The interpretation of results has to take into account that the experimental environment is populated with façades taken from a German city. This means explicit landmarks are labeled with signs in German, and some of the shops or restaurants are from chains well known to the German participants, but not necessarily to the other participants.

We found significant differences[1] ($p < 0.05$) between the "Germans" and the "non-Germans" in using explicit vs. implicit landmarks for both routes.

[1] T-Test is calculated with SPSS 16.

```
- At the red house take street so red house is on the left (the street
  with Haeagen-Dazs in the right)
- third intersection turn right, yellow house on corner, Lotto shop
  opposite yellow house
- next intersection turn left, house with balconies in the corner
- next intersection turn left, little corner restaurant
- next intersection at the big house turn right
- second intersection turn left, at the graffiti wall, there are a few
  shops in the street you turn into
- next intersection left, bright red house in the new street
- third intersection left, opposite the erdwaerme house
- next intersection right , into Heimstrasse
- next intersection left, at the white office style building with
  dark windows
- next right , at the Haeagen Dazs
- end at next intersection
```

Fig. 4. Sample route descriptions for the long route with references to landmarks

Unsurprisingly, German participants preferred explicit landmarks and non-Germans preferred implicit landmarks. However, there were no significant differences between these two groups in terms of the frequency with which they referred to landmarks.

4.2 Trajectory Analysis

We hypothesized that people refer to those façades in their descriptions that they had paid more attention to when walking through the environment. For example, participants might focus on façades that are visible longer along the route and refer to these in their descriptions. To test our hypothesis we tracked the position and heading (viewing direction) of each participant every millisecond during the experiments. By analyzing this data it could be computed how much time each participant spent looking at a particular façade. Figure 6 gives an example of such an analysis.

Each trajectory could be divided into two sets: the slow sections where positions were sampled with walking velocity, and the fast sections where the participants were running. For the analysis we concentrated on the slow parts, because we assumed that in running mode participants did not pay much attention to their environment and were not looking out for a landmark. In fact, every participant passed every intersection in the speed that was predefined as walking speed.

For every point of the slow sections we calculated a visibility area. The visibility area was defined by the heading of a participant at this point, the surrounding environment, a field of view of 60°, and a depth of view of 80 meters. In the computed visible area the set of visible façades was recorded, and then it was calculated how long a participant had seen a particular façade without interruption ($freq_{traj}$).

Fig. 5. The most prominent landmarks for both routes: on the left for the short route, on the right for the long route

We analyzed how often each façade was seen by all participants on average, and correlated this number ($freq_{traj}$) with the frequency this façade was referred to (freq_{human}). We found a significant correlation[2] ($p < 0.01$) between $freq_{traj}$ and the frequency of references to this façade for both the short and the long routes.

For a finer analysis we divided the set of all trajectories for the walking sections into a group where participants used landmarks ($freq_{withL}$), and a group where participants did not refer to landmarks ($freq_{noL}$). Following our hypothesis we found a significant correlation ($p < 0.01$) between $freq_{traj}$ and $freq_{human}$ in the group where participants referred to landmarks in their route descriptions, but there is no correlation in the group which did not use any landmarks.

4.3 Calculating Landmark Salience

To exclude landmarks far away or invisible to participants, only façades within a buffer of 80 meters along the selected routes were investigated for their landmarkness. Due to the consistent height of all buildings in the virtual environment and the façades' composite texture, the landmark salience was calculated using four attributes: their position relative to decision points (structural salience), their categories (semantic salience), their advance visibility and their color differences to their neighbors (visual salience).

The structural salience value $s_{structural}$ was computed according to [20]. Semantic salience $s_{semantic}$ was calculated from a manual assignment of a category salience according to [1], i.e., a value weighing prototypical physical size, prototypical proximity to road, prototypical ubiquity, and prototypical permanence. A list of used category salience measures is shown in Table 2. Only façades containing businesses have a semantic salience value $s_{semantic}$ larger than zero.

[2] Correlations are calculated using Spearman correlation and are done with SPSS 16. We used the Spearman rank correlation because one of our data sets is ordinal.

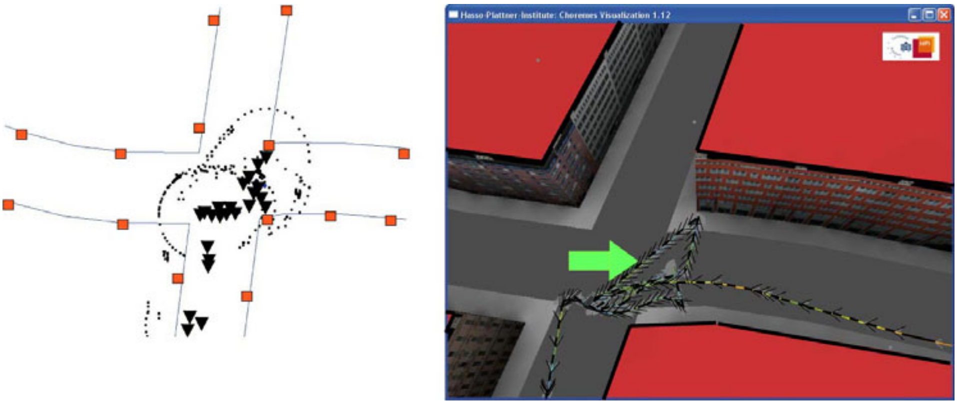


Fig. 6. Tracking a participant at an intersection. On the left side: squares are the ending point of a façade, black dots are the heading and triangles show the trace. On the right side: a visualization of the trajectory (dashed line) of the participant, with the opening angle of the heading. The arrow head to the direction of travel.

Table 1. Correlation between trajectory data (all, group using landmarks, group not using landmarks) and landmark references in the route descriptions (significant if $p <$ 0.01)

Route	$freq_{traj}$/ $freq_{human}$	$freq_{withL}$/ $freq_{human}$	$freq_{noL}$/ $freq_{human}$
short	0.16*	0.18*	not sig.
long	0.18*	0.17*	not sig.

Advance visibility s_{vAngle} of façades around decision points was calculated as their horizontal angle in the field of view from 50 meters before decision points, in proportion to the 60° of comfortable field of view of the human visual apparatus. For this computation the depth of view was set to 100 meters. To determine differences in color the individual façade images were used. Their median RGB color was transferred into an HSB color, and hue, saturation and brightness were weighted according to Nothegger et al. [24]. The adjusted color of each façade was calculated by:

$$color = 75\% * H + 20\% * S + 5\% * B \tag{1}$$

Then color salience $s_{colDiff}$ was determined as the standard deviation of color differences to the five nearest neighbors of each façades. The advance visibility s_{vAngle} can be zero for some façades as a result of this computation.

Also, it could be observed that participants described landmarks in different manners, such as by name or color. We were interested in how different attributes of landmark salience impact humans choice of landmarks. To examine the varying

Table 2. Category saliences as used in the experiment

Category	Salience
restaurant	0.886
business	0.682
hotel	0.773
street sign	0.2
shop	0.545
bank	0.705

impact, we used multiple linear regression (by *R* [3]) to find out how much the four attributes correlate with the landmarks participants referred to (w_1, w_2, w_3, w_4 are the coefficients):

$$c_{Salient} = w_1 * s_{structural} + w_2 * s_{semantic} + w_3 * s_{vAngle} + w_4 * s_{colDiff} \quad (2)$$

Table 3 shows various impact of the four salience attributes by correlating them to the frequency a landmark had been nominated by all participants for the short and long route. The *sum* field in the table is the Spearman rank correlation coefficients of façades with weighted salience and the human selected landmarks. It can be seen that structural salience and advance visibility dominate in their impact. In comparison, both semantic salience and color difference have less influence.

Table 3. Weights of four salience attributes suggested by multiple linear regression

Route	♯ F	structural	semantic	vAngle	colDiff	sum
short	256	1.01	0.41	4.92	0.14	0.30
long	429	2.64	0.18	2.15	-0.65	0.29

We additionally calculated the Spearman rank correlation coefficients for individual salience factors with the frequency of façades referenced (freq_{human}) (see Table 4). The results again show that the structural salience and advance visibility salience have stronger effects on the overall salience of landmarks in this environment for both routes.

Whether the nationality (Germans versus non-Germans) of the participants influenced the correlation of salience factors and the referenced landmarks, especially for semantic salience, was also tested. We only found a difference for semantic salience for the long route. There is a significant positive correlation (0.17) in the German group, but no significant correlation for the non-German group.

Finally we correlated each salience factor with the frequency each façade is seen along a route ($freq_{traj}$). For the short route a positive correlation ($p < 0.01$)

[3] http://www.r-project.org/

Table 4. Correlation of four salience attributes with $freq_{human}$. ♯ F shows the number of façades around the routes.

Route	♯ F	structural	semantic	vAngle	colDiff	sum
short	256	0.3	0.18	0.32	not sig.	0.27
long	429	0.47	not sig.	0.34	not sig.	0.28

Table 5. Correlation between trajectory data (all, group with landmarks, group with no landmarks) and the salience factors (significant if $p < 0.01$)

Route	structural/ $freq_{traj}$	structural/ $freq_{withL}$	structural/ $freq_{noL}$	vAngle/ $freq_{traj}$	vAngle/ $freq_{withL}$	vAngle/ $freq_{noL}$
short	0.16	0.21	not sig.	not sig.	not sig.	not sig.
long	not sig.	not sig.	not sig.	0.15	0.15	not sig.

existed only between structural salience and $freq_{traj}$. For the long route advance visibility salience correlated with $freq_{traj}$ ($p < 0.01$). Both correlations could also be found for the subset of the trajectories of the people who used landmarks ($freq_{withL}$).

5 Discussion

In this paper we take the first steps towards an evaluation of automatic landmark identification methods. We present an experiment in a virtual environment where participants were asked to write down route descriptions for a short and a long route. Although the environment is relatively sterile and populated only by textured box buildings, nearly two thirds of the participants referred to landmarks.

We correlated the frequency of landmarks referred to by participants with calculated salience values using four attributes of landmarks. These four salience attributes were structural salience, semantic salience, advance visibility salience, and visual salience based on color differences.

For both routes, the highest positive correlations occurred with advance visibility salience and structural salience. These results show that structural salience and advance visibility have higher impact on human selection of landmarksat least for the urban virtual environments tested.

Furthermore, these two factors were correlated with how often a façade was seen by all participants at walking speed. For the short route structural salience correlated positively with the visual attention determined from trajectory data, and for the long route we found a correlation with advance visibility.

With even higher impact, both structural salience and advance visibility were related to decision points. Analyses of the trajectory data proved that all participants were slower than average at each decision point, i.e., paid more attention to the environment.

One possible explanation why structural salience is important may be that this virtual environment is too homogeneous with respect to its visual appearance. The virtual environment is built of textured box buildings of equal height, and otherwise it is empty. A small set of façades and the constant building height may have downplayed the need for landmarks as well as the recognition factor of each façade, and may have led participants to concentrate on structural factors, such as structural salience and advance visibility salience.

Beyond this, we examined whether the nationalities of the participants influenced the selection of landmarks. Our urban virtual environment contains only German façades and so we divided our participants into a German and a non-German group. There were no differences between the two groups in the correlation analyses for almost all salience factors, except that for the long route the semantic salience factor correlated significantly to the referenced landmarks for the German group, but not for the non-German group.

The second question we investigated in this paper was to what extent the tracking data collected in the experiment provides information on how people select landmarks. Tracking data enabled us to calculate, for every position, which façade was in the field of view, and to sum up the visibility over all points. We correlated the time span each façade was looked at with the frequency of this façade in the route descriptions. We found a significant positive correlation between these two types of data, which suggests that tracking data provides useful information about which landmark will be referred to by people.

To elicit whether façades are looked at due to the movement pattern or because they will be referred to as landmarks, we compared the group of participants that did not refer to landmarks with the group that did. If we had found a correlation for both groups, façades would likely be looked at only because they happen to be at decision points where participants slowed down. Since we only found a correlation for the group that did refer to landmarks, it is more likely that attention is paid to those façades that are referenced. Thus, trajectory data can be a useful tool for getting additional information on what façades people will utilize as landmarks in virtual environments as well.

6 Conclusion and Outlook

In this paper we demonstrate that virtual environments can be utilized as a valuable tool to acquire more information about how people select landmarks. This paper contributes to the knowledge of automatic landmark identification theories, to a better understanding of the impact of individual salience factors, and to new methods of landmark identification, e.g., based on tracking what people experience in an environment.

Both experiment and interpretation can be further improved. During the analysis we identified some artifacts due to the experimental setup. For example, all participants showed a higher residence time at the start of the experiment, waiting for permission to begin.

We believe that we will get even higher correlations if we remove these artifacts. An analysis for each participants individual trajectory and its correlation to the individual written route descriptions may yield further interesting results.

Moreover, we are now interested in a finer analysis of the route descriptions according to inter-individual and inter-cultural differences. For this purpose we plan a follow-up experiment in Germany with participants of German and other nationalities.

We will also study more reasonable combinations and weightings of the landmark identification methods. In this paper we illustrated that individual factors in a salience formula do not need to be equally weighted. However, an adaptive weighting to environmental features or context would require more extensive testing across different virtual environments.

Acknowledgments

The authors acknowledge support by the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition, which is funded by the Deutsche Forschungsgemeinschaft (Peters), the Australian Research Councils Discovery Program (DP0878119, Wu), and the Go8/DAAD researcher exchange scheme (Peters). We also acknowledge the generous permission to use the virtual environment by the Hasso Plattner Institut, Potsdam. The authors are grateful for support by Tassilo Glander who helped us with the generation of the map information of the virtual environment and also with the visualization of the trajectories.

References

1. Duckham, M., Winter, S., Robinson, M.: Including landmarks in routing instructions. Submitted to an International Journal (2009)
2. Lynch, K.: The Image of the City. The MIT Press, Cambridge (1960)
3. Elias, B.: Extracting landmarks with data mining methods. In: Kuhn, W., Worboys, M.F., Timpf, S. (eds.) COSIT 2003. LNCS, vol. 2825, pp. 375–389. Springer, Heidelberg (2003)
4. Denis, M., Michon, P.E., Tom, A.: Assisting pedestrian wayfinding in urban settings: Why references to landmarks are crucial in direction-giving. In: Allen, G.L. (ed.) Applied Spatial Cognition: From Research to Cognitive Technology, pp. 25–51. Lawrence Erlbaum Associates, Mahwah (2007)
5. Egenhofer, M., Mark, D.: Naive geography. In: Frank, A.U., Kuhn, W. (eds.) COSIT 1995. LNCS, vol. 988, pp. 1–15. Springer, Heidelberg (1995)
6. Tenbrink, T.: Space, Time, and the Use of Language: An Investigation of Relationships. Mouton de Gruyter, Berlin (2007)
7. Clark, H.H.: Common ground (common experiences). In: Using Language, ch. 4. Cambridge University Press, Cambridge (1996)
8. Lovelace, K.L., Hegarty, M., Montello, D.R.: Elements of good route directions in familiar and unfamiliar environments. In: Freksa, C., Mark, D.M. (eds.) COSIT 1999. LNCS, vol. 1661, pp. 65–82. Springer, Heidelberg (1999)
9. May, A.J., Ross, T., Bayer, S.H.: Drivers' information requirements when navigating in an urban environment. The Journal of Navigation 56(01), 89–100 (2003)

10. Denis, M., Pazzaglia, F., Cornoldi, C., Bertolo, L.: Spatial discourse and navigation: An analysis of route directions in the city of venice. Applied Cognitive Psychology 13, 145–174 (1999)
11. Michon, P.E., Denis, M.: When and why are visual landmarks used in giving directions? In: Montello, D.R. (ed.) COSIT 2001. LNCS, vol. 2205, pp. 292–305. Springer, Heidelberg (2001)
12. Sorrows, M.E., Hirtle, S.C.: The nature of landmarks for real and electronic spaces. In: Freksa, C., Mark, D.M. (eds.) COSIT 1999. LNCS, vol. 1661, pp. 37–50. Springer, Heidelberg (1999)
13. Raubal, M., Winter, S.: Enriching wayfinding instructions with local landmarks. In: Egenhofer, M.J., Mark, D.M. (eds.) GIScience 2002. LNCS, vol. 2478, pp. 243–259. Springer, Heidelberg (2002)
14. Grabler, F., Agrawala, M., Sumner, R.W., Pauly, M.: Automatic generation of tourist maps. ACM Transactions on Graphics 27(3), Article 100, 1–11 (2008)
15. Tomko, M., Winter, S.: Recursive construction of granular route directions. Journal of Spatial Science 51(1), 101–115 (2006)
16. Winter, S.: Route adaptive selection of salient features. In: Kuhn, W., Worboys, M.F., Timpf, S. (eds.) COSIT 2003. LNCS, vol. 2825, pp. 349–361. Springer, Heidelberg (2003)
17. Brenner, C., Elias, B.: Extracting landmarks for car navigation systems using existing gis databases and laser scanning. In: PIA 2003, Munich, Germany, Part 3/W8 of ISPRS Archives, vol. XXXIV (2003)
18. Millonig, A., Schechtner, K.: Developing landmark-based pedestrian-navigation systems. IEEE Transactions on Intelligent Transportation Systems 8(1), 43–49 (2007)
19. Michon, P.E., Denis, M.: When and why are visual landmarks used in giving directions? In: Montello, D.R. (ed.) COSIT 2001. LNCS, vol. 2205, pp. 292–305. Springer, Heidelberg (2001)
20. Klippel, A., Winter, S.: Structural salience of landmarks for route directions. In: Cohn, A.G., Mark, D.M. (eds.) COSIT 2005. LNCS, vol. 3693, pp. 347–362. Springer, Heidelberg (2005)
21. Claramunt, C., Winter, S.: Structural salience of elements of the city. Environment and Planning B: Planning and Design 34, 1030–1050 (2007)
22. Anter, K.F.: What Colour is the Red House? Perceived colour of painted facades. Doctoral thesis, Royal Institute of Technology, Stockholm (2000)
23. Aziz, M.Z., Mertsching, B.: Color saliency and inhibition using static and dynamic scenes in region based visual attention. In: Paletta, L., Rome, E. (eds.) WAPCV 2007. LNCS (LNAI), vol. 4840, pp. 234–250. Springer, Heidelberg (2007)
24. Nothegger, C., Winter, S., Raubal, M.: Computation of the salience of features. Spatial Cognition and Computation 4(2), 113–136 (2004)
25. Nash, E.B., Edwards, G.W., Thompson, J.A., Barfield, W.: A review of presence and performance in virtual environments. International Journal of Human-computer Interaction 12(1), 1–41 (2000)
26. Mallot, H.A., Gillner, S.: Route navigating without place recognition: What is recognised in recognition-triggered responses? Perception 29, 43–55 (2000)
27. Steck, S.D., Mallot, H.A.: The role of global and local landmarks in virtual environment navigation. Presence 9, 69–83 (2000)
28. Ruddle, R.A., Payne, S., Jones, D.: Navigating building in"desk-top" virtual environments: Experimental investigation using extended navigational experience. Journal of Experimental Psychology 3, 143–159 (1997)

29. Peruch, P., Belingard, L., Thinus-Blanc, C.: Transfer of spatial knowledge from virtual to real environments. In: Freksa, C., Brauer, W., Habel, C., Wender, K.F. (eds.) Spatial Cognition 2000. LNCS (LNAI), vol. 1849, pp. 253–264. Springer, Heidelberg (2000)
30. Münzer, S., Stahl, C.: Providing individual route instructions for indoor wayfinding in complex, multi-level buildings. In: Probst, F., Keler, C. (eds.) GI-Days (2007)
31. Wallet, G., Sauzéon, H., Rodrigues, J., N'Kaoua, B.: Transfer of spatial knowledge from a virtual environment to reality: Impact of route complexity and subject's strategy on the exploration mode. Journal of Virtual Reality and Broadcasting 6(4) (2009)
32. Tversky, B., Lee, P.U.: How space structures language. In: Freksa, C., Habel, C., Wender, K.F. (eds.) Spatial Cognition 1998. LNCS (LNAI), vol. 1404, pp. 157–176. Springer, Heidelberg (1998)
33. Dale, R., Geldof, S., Prost, J.P.: Using natural language generation in automatic route description. Journal of Research and Practice in Information Technology 36(3), 23–39 (2004)
34. Ligozat, G.: From language to motion, and back: Generating and using route descriptions. In: Christodoulakis, D.N. (ed.) NLP 2000. LNCS (LNAI), vol. 1835, pp. 328–345. Springer, Heidelberg (2000)
35. Frank, A.U.: Pragmatic information content - how to measure the information in a route description. In: Duckham, M., Goodchild, M.F., Worboys, M.F. (eds.) Foundations in Geographic Information Science, pp. 47–68. Taylor & Francis, London (2003)
36. Glander, T., Peters, D., Döllner, J.: 3d wayfinding choremes: A cognitively motivated representation of route junctions in virtual environments. In: Proceedings of AGILE 2009 conference (2009) (Short Paper)

Men to the East and Women to the Right: Wayfinding with Verbal Route Instructions

Vanessa Joy A. Anacta and Angela Schwering

Institute for Geoinformatics, University of Muenster,
Weseler Strasse 253, 48151 Muenster, Germany
vanacta@yahoo.com, schwering@uni-muenster.de

Abstract. In this paper, we investigate the outdoor wayfinding performance of men and women in a shifting frame of reference with verbal route instructions given in German language. This study replicated the methodology of Ishikawa and Kiyomoto (2008) but investigated also the gender component. The participants were divided into absolute-relative (A-R) and relative-absolute (R-A) groups. They had to follow ten routes: The A-R group walked the first set of five routes with instructions given in absolute frame of reference which shifted to instructions in relative frame of reference. The R-A group, on the other hand, walked the first set of five routes with relative instructions and shifted to an absolute reference frame. In the experiment, the results showed that participants in both groups had difficulty following the absolute instructions wherein they had significantly more stops, more deviations and travelled longer off-route. The overall performance increase of participants who started with an absolute and shifted to a relative reference frame was higher than the performance decrease of participants who started with a relative and shifted to an absolute reference frame. In terms of gender, the wayfinding performance of both, men and women, was significantly better with instructions in relative than in absolute reference frame. Women made fewer stops, fewer deviations, and did not walk off the route as frequently as men. However, the gender effect was not significant.

Keywords: outdoor wayfinding, frame of reference, gender, verbal route instructions.

1 Introduction

Wayfinding is an interesting area of study as it investigates how people find their way in both familiar and unfamiliar environment. Differences lie on people's spatial abilities, with some finding it easier to locate a specific place compared to others. Clear route instructions are deemed important for people to follow any navigational task. It is necessary to place oneself in the situation of a person unfamiliar to the place and understand that people have their own preferences in finding their way. Some prefer cardinal directions (North, East, West, South) while others prefer a relative reference frame (right or left). Some prefer exact distances while others prefer approximation based on landmarks. Many researchers in various fields such as geography, psychology, architecture, engineering, and computer science have investigated such navigational differences and how to make wayfinding easier for people.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 70–84, 2010.

This research attempts to replicate the methodology used by Ishikawa and Kiyomoto (2008) to study the effect of 'shifting reference frames' on wayfinding performance of German speakers. The shifting of reference frame makes the evaluation of an individual's level of spatial knowledge possible. In contrast to Ishikawa and Kiyomoto's study, this research investigates gender differences in preference of frame of reference. Gender differences are analyzed based on the outdoor wayfinding performances measures used which include the number of stops, number of deviations and the off-route distances in a specific frame of reference. This research examines which frame of reference men and women are most comfortable with, and attempts to assess whether men perform better in absolute frame of reference, while women perform better in relative directions. The cultural differences between German and Japanese outdoor wayfinding allows for an interesting comparative study and a comparison to Ishikawa and Kiyomoto's results. Further research is suggested to gain more insights on cultural differences in wayfinding. This could also be used in assessing how people orient themselves in space and how they normally refer to spatial models in reaching a specific destination.

This study posed the following hypotheses: 1) Men and women have difficulty following wayfinding instructions in a shifting frame of reference, but more so in the absolute frame; 2) Men perform better than women in outdoor wayfinding following verbal instructions in absolute directions; and 3) Women mostly have difficulty following absolute instructions rather than relative frame of reference.

The remainder of this paper is structured as follows: Section 2 gives an overview of related work regarding different reference frames and gender differences in wayfinding. In Section 3, we describe the methodology of our experiment. Section 4 gives the results with respect to the different reference frames and gender differences and section 5 discusses the results. The conclusions are drawn in section 6.

2 Related Work

Levinson (1996) described three different frames of reference: intrinsic, relative, and absolute. In the intrinsic reference frame, the orientation is defined with respect to the reference object; in the relative reference frame, orientation is defined with respect to the viewer; and in the absolute reference frame, orientation is defined with respect to fixed bearings such as cardinal directions.

Wayfinding directions are usually given in either relative or absolute frame of reference. In the following navigation task, people's preferences vary on which is easier to follow. Several factors such as the time spent and the number of stops are taken into account while following directions.

Golledge et al. (1992) emphasized the need to understand how people acquire knowledge of the environment. Route learning has captured interest of several researchers from various disciplines, mostly in geography. According to Golledge et al, there is superiority of survey learning in unfamiliar environment at varying geographic background with gender differences. Lawton (2001) looked at how the geography of the place may affect how people would give directions such as the case of the Midwest/West region in the US whose inhabitants prefer giving cardinal directions, which might be attributed to their grid-like pattern road network.

Lovelace and Montello (1999) investigated various ways of assessing the quality of route directions through the inclusion of landmarks, turns, segments, and descriptive information. These are considered helpful measures to guide people in any wayfinding task.

Researchers have also noted cultural differences; for instance, Japanese speakers prefer relative frame of reference (Ishikawa and Kiyomoto, 2008). Ishikawa and Kiyomoto looked at how Japanese students fared when there is a shift from relative to absolute direction or vice versa. Also of equal interest are other ethnic groups such as Tenejapan Tzeltal that use only absolute directions while some Mayan cultures such as the Mopan prefers an intrinsic frame of reference (Levinson, 1996).

Several studies have been conducted regarding how men and women use their spatial abilities in navigating. Many experiments aimed to identify how gender or sex differences affect such an activity. Collucia and Louse's (2004) review on gender differences in spatial orientation revealed mixed results of spatial differences. However, it mostly showed the rarity of female superiority in many spatial tasks, attributable to biological explanations and environmental factors. According to Collucia and Louse, gender differences in terms of spatial orientation may be explained through evolutionistic approach, individual strategies and personality factors. Ward (1986) showed that people mostly rely on listing landmarks and turns while giving directions whereby men would use more absolute indicators than women. Although many studies showed similar result, Weiss et al. (2003) regarded the concept of spatial ability and spatial cognition as vague.

It is often considered a stereotype that women have more difficulty in any wayfinding task than men (O'Laughlin and Brubaker, 1998). There is a tendency for women to even think that they have a poor sense of direction and they sometimes cannot follow direction, especially in an absolute frame of reference. However, the results are diverse, with some saying that men are better than women while others say there is no gender difference. Eals and Silverman (1994) highlighted the fact that there is a female advantage in recalling spatial objects. Montello et al (1999) further emphasized the superiority of women in object location and that they make fewer mistakes in recalling landmarks from their campus route experiment. Men, on the other hand, showed better spatial ability by using metric distances and cardinal directions when thinking of environmental space. Montello et al. pointed out that it is wrong to assume that males are generally better than females in terms of spatial abilities because they vary at different levels. Kim et al (2007) showed a similar result whereby female subjects responded faster in a 2-D matrix navigation task than males when landmark instructions were provided. The differences is quite interesting as it follows the idea of women being more confident in relative direction-related tasks while men are more on absolute directions. However, Iachini et al (2005) found no differences between men and women in object recognition and in recalling the spatial layout of a place.

There is male advantage in angular judgment and travel distance elimination (Holding and Holding, 1988). It shows that most women tend to show bias towards underestimation and that they were guessing in some spatial tasks. In a neuropsychological test conducted, they found out that men are better on visual-spatial tasks such as spatial orientation, mechanical abilities and mathematics while women outperformed men on most verbal tests (Weiss, et al, 2003). Although, there is a decreasing gender difference in mental rotations test over time (Masters and Sanders, 1993; Colom, et al., 1999).

Scholl et al. (2000) showed that men are better in orientation strategy (which refers to cardinal directions) while women use route strategy referring mostly to landmarks. The authors observed that women prefer route strategies when asking for directions and tracking the distances when going through the routes in an unfamiliar environment, while men use cardinal directions more often for orientation. In the study of Silverman and Choi (2006), results revealed that men outperformed women in dynamic navigation following Euclidean instructions. Women mostly use topographical over Euclidean navigational strategies. Schmitz (1997) studied how gender played a role on some German students in wayfinding in a three-dimensional maze. The results showed that girls developed higher anxiety and fear than boys when going through the maze. It also turned out that those subjects who were slow in the experiment and scored high in anxiety and fear tend to recall more landmarks than those who were less nervous. Women usually tend to experience higher spatial anxiety over men (Lawton, 1994), which also appeared in a cross cultural study between Hungarians and Americans (Lawton and Kallai, 2001).

There are an increasing number of studies showing no gender differences in spatial cognition specifically in learning a spatial skill. Spence et al. (2009) trained selected participants in learning a new video game. Such training method proved that women could equally acquire a basic spatial skill like men. Hund and Minarik (2006) stated in their study that men and women showed they were both fast and accurate when navigating based on cardinal directions than in landmark directions.

3 Outdoor Wayfinding Experiment: Method

3.1 Participants

Twenty-four German students in the undergraduate and graduate levels unfamiliar to the study area participated in the experiment. The group was composed of 12 men and 12 women who were not compensated for participating. The age range was 19 to 30 years old with mean age of 23.88 years. The participants were all geosciences students except for one political science student.

3.2 Study Area

The study site was a residential area in Münster, Germany, which was purposely chosen because of its unfamiliarity to the participants (see Fig.1). Route lengths and number of turns were taken into consideration in choosing a study area, comparable to that of Ishikawa and Kiyomoto's experiment.

3.3 Procedure

Snowball sampling method was used in gathering participants. The participants were divided into Group A-R (Absolute-Relative) and Group R-A (Relative-Absolute), each group with equal distribution of men and women. The A-R group walked the first set of five routes (Route A) in absolute directions and then shifted to relative directions on the second set of route segment (Route B). The R-A group, on the other hand, walked Route A given in relative directions first and then shifted to absolute directions in Route B.

Before the participants started the experiment, they were asked to fill out the Santa Barbara Sense of Direction Scale Questionnaire which consists of 15 questions, which were then used in correlating to each individual's performance in the wayfinding task. Seven questions were stated in a positive statement such as: "My "sense of direction" is very good," while eight questions were negatively phrased, i.e.: "I don't remember routes very well while riding as a passenger in a car." The questionnaire was used to examine the spatial and navigational abilities, preferences and experiences of each individual.

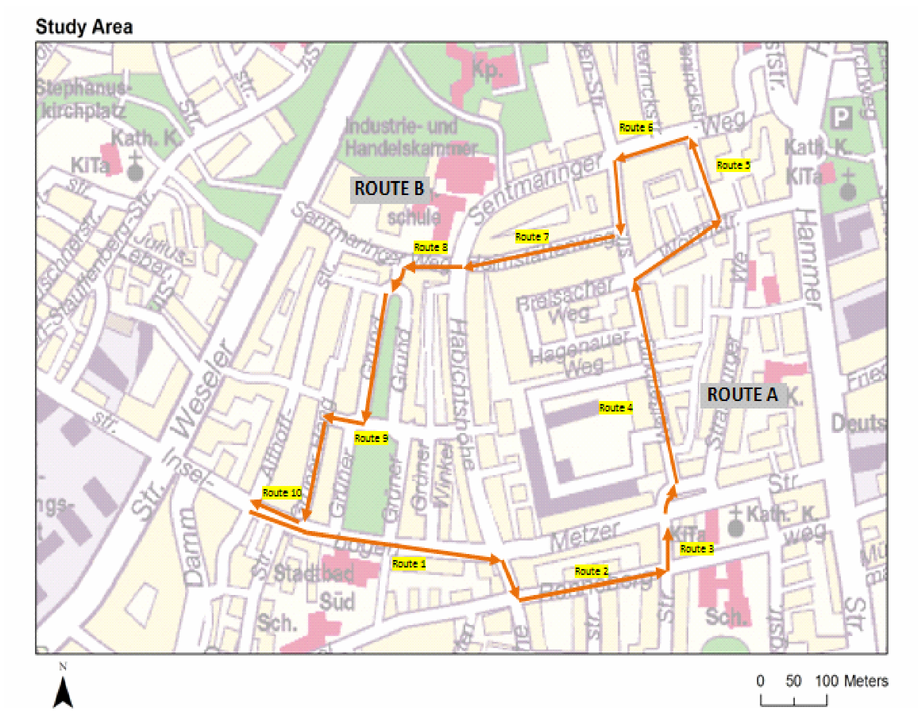


Fig. 1. The study area with the 10 routes. Route A comprises 1-5 routes while Route B is composed of routes 6-10.

The entire procedure was explained to the participants, which includes taking the video while conducting the experiment. The cardinal directions were explained. Before proceeding to the starting point, each participant was asked to point where north is. The participant was then led to the starting point where the first task was given. The participant was followed by the experimenter without making any conversation. Whenever the participant take the wrong route, the experimenter leads him/her to the right direction once s/he takes twice the expected time to travel the entire route. This was also used in gauging the off-route distance they travelled which doubled the actual distance of the route. The whole wayfinding experiment was videotaped to record the time spent, deviations of the route and the number of stops made. Spending 15 seconds or longer on a particular spot counts as one stop.

3.4 Verbal Route Instructions

Each route had a distance which ranged between 62-400 meters long. The entire route was 2.3 km. A participant had to find a specific goal before proceeding to the next route. For relative reference frame, egocentric representations were used for directions. For the absolute part, cardinal directions and distances in meters were used. The same landmarks as sub-goals were used for both absolute and relative frames of reference for each route.

In the verbal route directions, the participant was given an 11x15 cm instruction card where the routes were described in either relative or absolute frame. An example of a relative route instruction is Route 3 written as: "Turn left from the pharmacy and walk straight. You see the Ulf Import driving school to your right. Cross the street, you see the Haus Niemann restaurant **[goal]** to your left. You see Johanniter-Akademie guest house in front of you. You also see Schlecker store and a church to your right." (Translated from German[1]).

The absolute instruction version of the same route reads, "Walk 70 meters north. You see the Ulf Import driving school to the east. Cross the east-west running road, you see a Haus Niemann restaurant **[goal]** to the west. You see Johanniter-Akademie guest house to the north. You also see Schlecker store to the east and a church to the ESE." (Translated from German[2]).

4 Results

4.1 Groups A-R and R-A on Switching Frame of Reference

The results were analyzed in SPSS using Mixed ANOVA with the Route sets (Route A and Route B) as within-subject factor and the Group (Absolute-Relative and Relative-Absolute) as between-subject factor.

4.1.1 Number of Stops

A mixed ANOVA showed that the main effect of Group was significant, F (1, 22) = 4.25, p =0.05. Hence, there was a significant difference with the number of stops each group made. There was a significant effect of Route set (stops) and Route set (stops) x

[1] The original instructions in German read as follows "*Bitte biegen Sie an der Apotheke links ab und laufen geradeaus. Sie sehen die Ulf Import Fahrschule zu Ihrer Rechten. Überqueren Sie eine Straße. Zu Ihrer Linken sehen Sie das Restaurant Haus Niemann [**Ziel**]. Vor Ihnen sehen Sie die Johanniter Akademie Gästehaus. Außerdem sehen Sie eine Kirche und einen Schlecker Supermarkt zu Ihrer Rechten.*"

[2] The original German instructions read as follows: "*Bitte laufen Sie 70 Meter in Richtung Norden. Sie sehen die Ulf Import Fahrschule im Osten. Überqueren Sie nun eine Ost-West verlaufende Straße. Sie sehen das Restaurant Haus Niemann [**Ziel**] im Westen. Sie sehen die Johanniter Akademie Gästehaus im Norden. Außerdem sehen Sie eine Kirche im OSO und einen Schlecker Supermarkt im Osten.*"

Group interaction with F (1, 22) = 7.615, p < 0.05 and F (1, 22) = 10.36, p < 0.01, respectively. This means that participants made more stops on one set of five routes which was the absolute frame of reference.The number of stops made by the A-R group in the absolute direction is (on average per participant) 2.7, but only 0.5 in the relative frame. The RA group, on the other hand, made 0.6 stops in the relative and 0.8 in the absolute frame of reference.

Fig. 2 shows the plotted result for A-R group, indicating that the participants made more stops in the first set of routes following absolute instructions, with the number of stops decreasing significantly in the second set with relative instructions (from point P1 to P2). For RA group, there is a minimal increase of the number of stops in the second set of routes in terms of the absolute directions (from point P4 to P3 in Fig. 2).

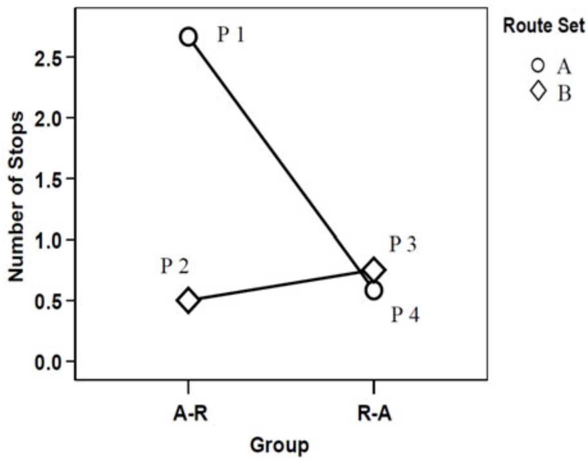


Fig. 2. The plotted graph shows the number of stops for A-R and R-A groups. P1 represents Route A (absolute) and P2 Route B (relative) for group A-R. P3 represents Route B (absolute) and P4 represents Route A (relative) for group R-A.

4.1.2 Number of Deviations

All deviations off the route were recorded in each route. Also, the off-route distances were taken into account to determine how far each participant deviated from a particular route. Similar to the number of stops, mixed ANOVA was used for analysis. It showed an equal number of deviations for both R-A and A-R group wherein each group had 12 deviations in the entire experiment.

The main effect of route set was not significant F (1, 22) = 1.165, p > 0.05. The number of deviations for the A-R group following absolute directions was (on average per participant) 0.8 and only 0.2 deviations in the relative frame. The R-A group participants made 0.4 deviations in the relative direction and 0.6 deviations in the absolute frame.

Fig. 3 shows that A-R group had many deviations when they began the route following the absolute frame instructions (Route A) and fewer deviations when it shifted to relative directions (Route B). Participants of the R-A group, on the other hand, made less deviations in the relative frame (Route A) but more deviations when the absolute direction (Route B) was given.

4.1.3 Off-Route Distances

There is a significant effect of off-route distance and group interaction, $F(1, 22) = 4.96$, $p < 0.05$. For A-R group in the absolute frame, participants travelled 99 m (on average per participant), but in the relative frame, they only went off-route for 11 m. The R-A group, on the other hand, traveled 24 m off-route in the relative frame while 50 m in the absolute direction.

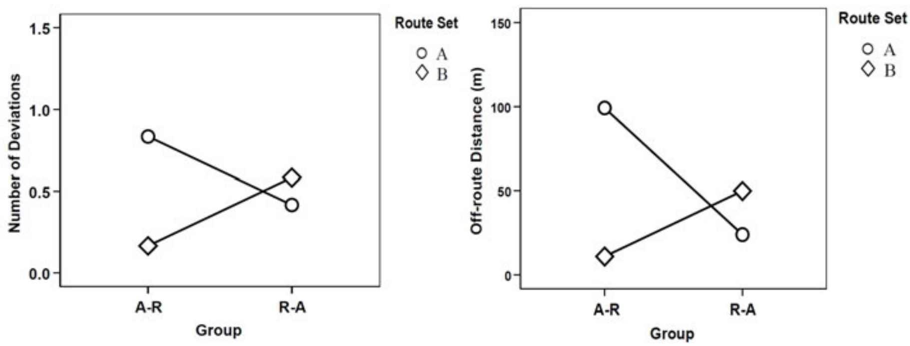


Fig. 3. The plotted graph shows the number of deviations and off-route distance of A-R and R-A groups. Both figures show a decrease of number of deviations and off-route distance for the A-R group when shifting to the relative reference frame on route B. In the R-A group we can see an increase in number of deviations and off-route distance for the R-A group when shifting to the absolute reference frame on route B.

4.1.4 Travel Time and Walking Speed

The Mixed ANOVA shows that the main effect of route set was significant as well as the route set for travel time and group interaction, $F(1, 22) = 47.65$, $p < 0.001$ and $F(1, 22) = 15.52$, $p < 0.01$, respectively. Hence, this means that there are differences in the travel time in the route set of both absolute and relative frames. For A-R group, participant's travel time (on average per participant) was 26min on route A with the absolute directions and 15 min on route B with the relative directions. For R-A group, the participant's travel time was 20 min on route A with relative directions and 17 min on route B with absolute directions.

The walking speed was recorded in terms of steps per minute of each participant in every route. It was averaged for both Route A and Route B to examine whether there has been a change in their speed. It showed that there is no significant effect in terms of walking speed for both groups, $F(1, 22) = 2.78$, $p > 0.05$. This means that participants maintained their normal walking speed even when a shift to another frame of reference was given.

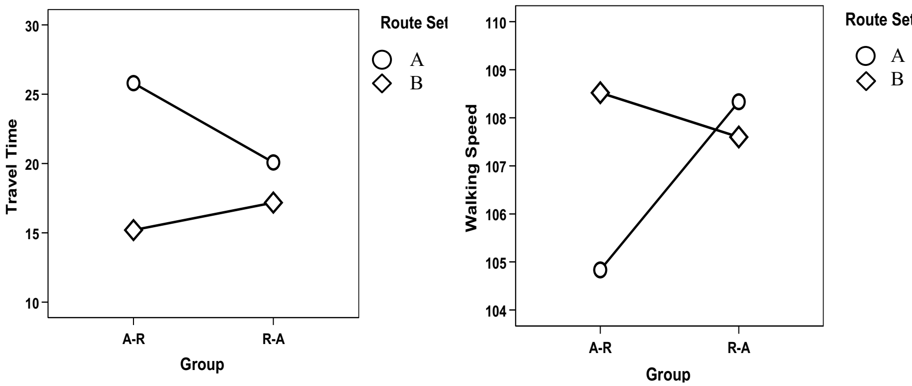


Fig. 4. Travel time and walking speed for A-R and R-A groups

4.2 Men and Women's Performance on the Frames of Reference

Mixed ANOVA was also used in analyzing gender performance in outdoor wayfinding with the Frame of Reference set (Absolute and Relative) as the within-subject factor and the Gender (Men and Women) as the between-subject factor.

4.2.1 Number of Stops

The main effect of frame of reference set was significant, $F(1, 22) = 7.42$, $p < 0.01$, showing a significant difference on how men and women perform in absolute and relative frame of references. Fig. 5 shows that there were fewer stops in the relative directions made by both gender, and men made more stops than women.

The number of stops made by women in absolute and relative frames of reference is on average per person 1.5 and 0.3, respectively. Men meanwhile, made 1.9 stops following absolute directions and 0.8 with relative directions.

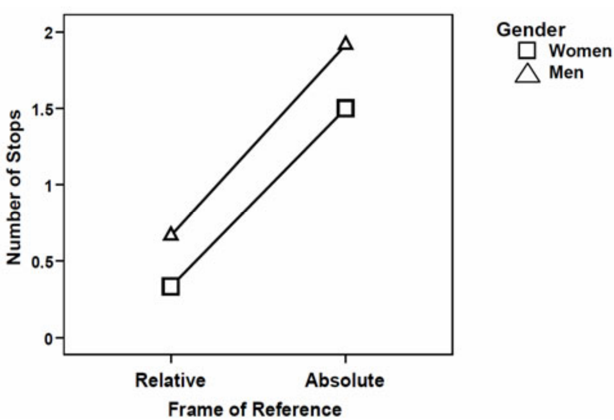


Fig. 5. The plotted graph shows the number of stops of men and women in absolute and relative frames of reference

4.2.2 Number of Deviations

There was no significant effect of reference frame for gender, $F(1,22) = 3.14$, $p > 0.05$. Although, it showed in the result that the absolute frame of reference was a bit difficult for participants to follow. Fig. 6 shows that there had been more deviations in the absolute as compared to the relative frame of reference. It also shows that men made more deviations than women. The number of deviations men incurred for the relative frame is 0.33 while in the absolute frame, they made 0.92 deviations. For women, they made 0.25 deviations in the relative frame and 0.5 deviations following the absolute direction.

4.2.3 Off-Route Distances

In terms of the off-route distance, the main effect of frame of reference set was significant, $F(1, 22) = 4.86$, $p < 0.05$. Both men and women walked longer off-route distances in the absolute frame of reference. However, as Fig. 6 shows, men walked longer distances when they went off-route in both frames of reference. Men walked on average 103m off-route following absolute directions and 20m off-route following relative directions. Women walked on average 46m off-route in absolute directions and 15m off-route following using relative directions.

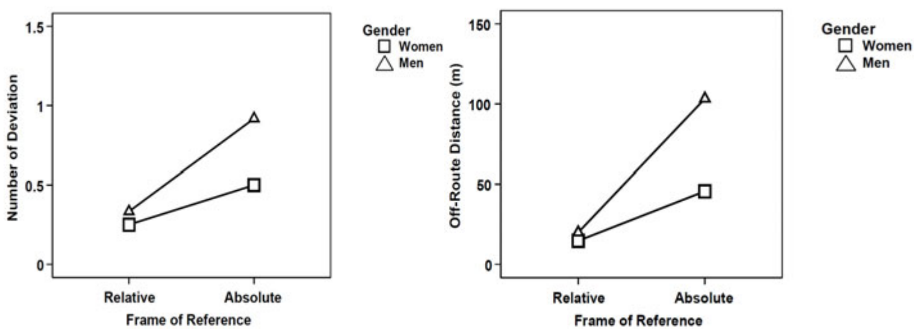


Fig. 6. The plotted graph shows the number of deviations and off route distance of men and women made in absolute and relative frames of reference

4.2.4 Travel Time and Walking Speed

For the time each group spent on the experiment, there was a significant effect of the frame of reference set, $F(1,22) = 5.03$, $p < 0.05$, indicating that both men and women spent longer time following the absolute frame of reference. Women spent on average 21min with absolute and 18min with relative directions. Men spent on average 22min with absolute and 17min with relative directions.

There was likewise a significant effect of the frame of reference set in terms of walking speed, $F(1,22) = 4.86$, $p < 0.05$. Both men and women walked faster in the relative frame than in the absolute. It was also observed that women's walking speed was slightly faster than that of men. This might be the effect of the changing weather condition wherein most women participated in the later part of the experiment when it started to snow. Walking speed of women was on average 108 (steps per minute) in the absolute and 111 in the relative reference frame and men's walking speed was 104 in absolute and 106 in relative reference frame.

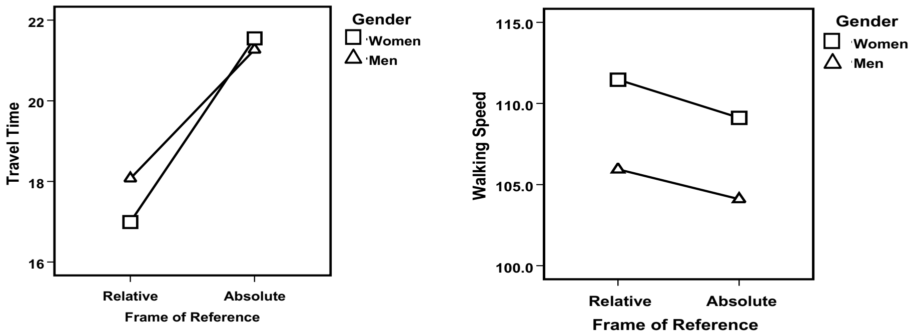


Fig. 7. Travel time and walking speed of men and women

4.3 Correlation with the Sense of Direction Scale

The Santa Barbara Sense-of-Direction questionnaire is an indicator of how well the participant judged his/her spatial and navigational abilities, preferences, and experiences. The questions ranged from assessing how poor his/her memory is and how well s/he enjoys reading maps and giving directions.

The mean was calculated for each participant. The negative questions were translated into positively formed statements so that the high numbers mean that participants rated their own sense of direction (SOD) as good. The result was then correlated with the number of stops, deviations, and off-route distances.

Using a Pearson correlation measure, it showed that there was a negative correlation with the sense of direction scale and the three wayfinding performance measures in the first set of routes for absolute frame of reference. This means that participants with better SOD tend to make fewer stops, less deviations, and walked off the route less frequently. The result was, stops ($r = -0.60$, $p < 0.05$), deviations ($r = -0.21$, $p > 0.05$), and off-route ($r = -0.02$, $p > 0.05$). However, the correlation was only significant for the number of stops.

The correlation result of men and women revealed that there was a negative correlation for all the performance measures along with the SOD of women although not significant; stops ($r = -0.47$, $p > 0.05$), deviations ($r = -0.31$, $p > 0.05$), and off-route distance ($r = -0.27$, $p > 0.05$). However, for men, it was observed that there was only a negative correlation for the number of stops but a positive correlation for deviations and off-route distances; stops ($r = -0.22$, $p > 0.05$), deviations ($r = 0.43$, $p > 0.05$), and off-route ($r = 0.29$, $p > 0.05$). Also, the result did not show any significant effect on gender difference. A positive correlation means that those participants, who judged their SOD higher, tend to make more stops, more deviations, and make longer off-route distances.

5 Discussion

5.1 Shift and Comparison of Results to Ishikawa and Kiyomoto (2008)

The findings of this study showed that there are differences on how people perform outdoor wayfinding experiment depending on which frame of reference was used.

Generally, subjects performed better when instructions were given in a relative reference frame and there were more stops and deviations when instructions were given in absolute directions, particularly when participants started off with instructions in the absolute reference frame. When participants started with instructions in the relative reference frame, their performance got worse when shifting to the absolute reference frame. But the participants were still able to adapt to their non-preferred (absolute) frame of reference. In this experiment, it is observed that the A-R group participants found it difficult to follow absolute instructions first, considering the number of stops and deviations they made. When shifting to the preferred frame of reference, the relative reference frame, a notable improvement in their wayfinding performance was seen. In the R-A group, participants shifted from the preferred relative reference frame to the non-preferred, absolute frame of reference. In this group, the wayfinding performance was also better in relative than in absolute reference frame, but the differences was not as big as in the A-R group. The experiment showed that shifting to a non-preferred reference frame was not as difficult for the participants who started in the preferred frame of reference.

This research followed the methodology of Ishikawa and Kiyomoto (2008), examining how people adapt to switching frames of reference following verbal route instructions. Comparing the results to their study, it also shows that German speakers just like the Japanese, prefer relative frame of reference. In Table 1, both Japanese and German participants showed significant effects for the interaction of Route and Group in the number of stops, off-route distance, and travel time.

In terms of the walking speed, the same result was attained comparing with Ishikawa and Kiyomoto's experiment wherein there was no difference in the participant's speed even when the route instructions were changed. Participants were still able to maintain their normal walking speed when a shift of reference frame was made.

Table 1. Comparison of German and Japanese Participants

Wayfinding Measures	**German Participants**	**Japanese Participants** (Ishikawa and Kiyomoto, 2008)
Number of Stops	sig. for Route x Group	sig. for Route x Group*
Deviations	not significant	sig. for Route x Group*
Off-route Distance	sig. for Route x Group	sig. for Route x Group*
Travel Time	sig. for Route x Group	sig. for Route x Group**
Walking Speed	not significant	not significant

* Due to difficulties on Route 2, it was excluded in Ishikawa and Kiyomoto
** only route B

Similarly to Ishikawa and Kiyomoto's results, the correlation between the SOD as assessed by the participants corresponding to their wayfinding performance was negative. This means that those who judged themselves having good SOD made less deviations, few stops and did not walk off the route frequently. However, the correlation in this study was only significant for the number of stops. In the case of off-route distances, Ishikawa and Kiyomoto found a stronger negative correlation, while the present study showed very low negative correlation.

5.2 Gender Differences on the Frame of Reference

Recalling the hypotheses set in the study that (1) men perform better than women in outdoor wayfinding task where they had to follow verbal instructions in absolute directions and (2) women oftentimes have difficulty following absolute instructions compared with instructions in relative frame of reference; both hypotheses could not be verified in this study. The results even showed a tendency for women to perform better.

Based on the wayfinding performance measures set, the results showed that women made less stops, deviations and went off- route less frequently. However, the effect size of the gender differences is not big enough to show significant results with 24 subjects of this experiment. Studies stating that men usually perform better than women in many spatial activities especially in wayfinding (O'Laughlin and Brubaker, 1998; Lawton, 1994) could not be supported. In terms of the frame of reference, this experiment revealed that both genders perform significantly better with instructions in relative reference frame than with absolute reference frame. Several researchers have shown that men are better in following absolute frame of reference (Holding and Holding, 1988; Scholl et al, 2000; Silverman and Choi, 2006;). However, the experiment showed that women were able to follow absolute directions better than men following verbal instructions. An interesting observation is that when instruction shifted from relative to absolute frame of reference, the female participants did not make deviations. It was observed that women tend to adapt more easily to changing route instructions than men.

Holding and Holding (1998) stated in their study that women tend to show bias towards underestimation and that they were guessing in some spatial tasks, but in this research, women were able to estimate their distances by not walking longer off route as compared to men in this case, who tend to overestimate their distances when they went off the route.

In terms of the correlation of sense-of-direction report and wayfinding performance, the study did not reveal any significant correlations. The non-significant correlation indicated that women were able to evaluate their spatial skills better than men.

6 Conclusions and Future Work

This study presented interesting results on varying spatial abilities of men and women in wayfinding with different reference frames. This research contributes to existing literature on analyzing wayfinding and gender differences. The shifting frame of reference is seen as important to determine whether people follow navigational instructions in relative or absolute frame of reference more efficiently.

Our study showed that people perform significantly better in wayfinding if the route instructions are given in a relative reference frame compared to an absolute reference frame. This is independent of gender whereby men and women equally showed this significant preference for relative instructions.

Furthermore, we found a non-significant tendency that women outperform men in following relative and absolute directions. This contradicts to other studies which see women as poor in wayfinding task, particularly with cardinal directions. Our findings are based on several indicators to evaluate wayfinding performance. Women made fewer mistakes regarding the number of stops, deviations, and the off-route distance.

Analyzing the reference frame shift, we found that the overall performance increase of participants who started with an absolute and shifted to a relative reference frame was higher than the performance decrease of participants who started with a relative and shifted to an absolute reference frame. It seems to be easier to switch from a preferred to a non-preferred reference frame than starting with the non-preferred reference frame immediately. On the other hand, people who started with the non-preferred reference frame seem to feel more comfortable when switching to the preferred reference frame than when they started with it immediately.

While the experiment with a group of 24 participants showed significant effects with respect to the preferred reference frame, the gender effect was not significant. Future work needs to address this aspect and increase the sample size to test whether the female priority in number of stops, number of deviations and off-route distance is significant. Moreover, future work will examine the navigational ability of people across culture.

Acknowledgements

We would like to thank the European Commission for the scholarship granted to the first author on MSc. in Geospatial Technologies under the Erasmus Mundus Programme. We acknowledge the valuable insights of Professor Toru Ishikawa during his visit at the institute. Furthermore we thank the four anonymous reviewers for their helpful comments and the participants for taking part in our experiment.

References

Colom, R., Quiroga, M.A., Juan-Espinosa, M.: Are cognitive sex differences disappearing? Evidence from Spanish populations. Personality and Individual Differences (27), 1189–1195 (1999)

Coluccia, E., Louse, G.: Gender differences in spatial orientation: A Review. Journal of Environmental Psychology 24, 329–340 (2004)

Eals, M., Silverman, I.: The Hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in Recall of object arrays. Ethology and Sociobiology 15, 105 (1994)

Golledge, R.G., Dougherty, V., Bell, S.: Survey versus Route-Based Wayfinding in Unfamiliar Environments. Working Paper UCTC No. 214. Paper presented at AAG Conference. University of California Transportation Center (1993)

Golledge, R.G., Gale, N., Pellegrino, J.W., Doherty, S.: Spatial Knowledge Aquisition by Children: Route learning and relational distances. Annals of the Association of American Geographers 82(2), 223–244 (1992)

Holding, C.S., Holding, D.H.: Acquisition of Route Network Knowledge by Males and Females. The Journal of General Psychology 116, 29–41 (1988)

Hund, A., Minarik, J.: Getting from here to there: Spatial Anxiety, Wayfinding Strategies, Direction Type, and Wayfinding Efficiency. Spatial Cognition and Computation 6(3), 179–201 (2006)

Iachini, T., Sergi, I., Ruggiero, G., Gnisci, A.: Gender differences in object location memory in a real three-dimensional environment. Brain and Cognition 59, 52–59 (2005)

Ishikawa, T., Kiyomoto, M.: Turn to the Left or to the West: Verbal Navigational Directions in Relative and Absolute Frames of Reference. In: Cova, T.J., Miller, H.J., Beard, K., Frank, A.U., Goodchild, M.F. (eds.) GIScience 2008. LNCS, vol. 5266, pp. 119–132. Springer, Heidelberg (2008)

Kim, B., Lee, S., Lee, J.: Gender Differences in Spatial Navigation. World Academy of Science, Engineering and Technology 31 (2007)

Lawton, C.A., Kallai, J.: Gender Differences in Wayfinding Strategies and Anxiety about wayfinding: A cross-cultural comparison. Sex Roles 47(9/10), 389–401 (2002)

Lawton, C.A.: Gender and Regional Differences in Spatial Referents used in Direction giving. Sex Roles 44(5/6), 321–337 (2001)

Levinson, S.C.: Language and Space. Annual Reviews on Anthropology 25, 353–382 (1996)

Lovelace, K.L., Hegarty, M., Montello, D.: Elements of Good Route Directions in Familiar and Unfamiliar Environments. In: Freksa, C., Mark, D.M. (eds.) COSIT 1999. LNCS, vol. 1661, pp. 65–82. Springer, Heidelberg (1999)

Masters, M.S., Sanders, B.: Is the Gender Difference in Mental Rotation Disappearing? Behaviour Genetics 23(4) (1993)

Montello, D., Lovelace, K., Golledge, R., Self, C.: Sex-related differences and similarities in Geographic and Environmental Spatial Abilities. Annals of Association of American Geographers 89(3), 515–534 (1999)

O'Laughlin, E.O., Brubaker, B.S.: Use of Landmarks in Cognitive Mapping: Gender Differences in Self Report versus Performance. Person, Individual Difference 24(5), 595–601 (1998); Santa Barbara Sense of Direction Scale (SBSOD), `http://www.psych.ucsb.edu/~hegarty/instruments/sbsod.pdf` (Last accessed on September 10, 2009)

Schmitz, S.: Gender-related strategies in environmental development: Effects of anxiety on wayfinding in and representation of a three-dimensional maze. Journal of Environmental Psychology 17, 215–228 (1997)

Scholl, M.J., Acacio, J.C., Makar, R.O., Leon, C.: The Relation of Sex and Sense of Direction to Spatial Orientation in an unfamiliar environment. Journal of Environmental Psychology 20, 17–28 (2000)

Silverman, I., Choi, J.: Non-Euclidean Navigational Strategies of Women: Compensatory Response or Evolved Dimorphism? Evolutionary Psychology 4, 75–84 (2006)

Spence, I., Yu, J.J., Feng, J., Marshman, J.: Women match men when learning a spatial skill. Journal of Experimental Psychology: Learning, Memory, and Cognition 35(4), 1097–1103 (2009)

Ward, S.L., Newcombe, N., Overton, W.F.: Turn Left at the Church, or Three Miles North: A Study of Direction Giving and Sex Differences. Environment and Behavior 18(2), 192–213 (1986)

Weiss, E.M., et al.: Sex Differences in cognitive functions. Personality and Individual Differences 35, 863–875 (2003)

Do All Science Disciplines Rely on Spatial Abilities? Preliminary Evidence from Self-report Questionnaires

Mary Hegarty[1], Raymond D. Crookes[2], Drew Dara-Abrams[1], and Thomas F. Shipley[2]

[1] University of California, Santa Barbara
Santa Barbara, CA 93106-9660 USA
{hegarty,dara-abrams}@psych.ucsb.edu
[2] Temple University,
Philadelphia, PA 19144 USA
{rcrookes,tshipley}@temple.edu

Abstract. Spatial abilities are correlated with success in science. However, research on this topic has not focused on possible differences in the spatial demands of different scientific fields. Furthermore, there is a partial dissociation between spatial abilities involved in dealing with the space of environments (large-scale spatial abilities) and those involved in dealing with the space of objects (small-scale spatial abilities). We used on-line self-report measures to collect preliminary data on the spatial abilities of scientists in different fields, as well as humanists and individuals in professional fields. Geoscientists had the highest self-report ratings of both environmental and small-scale spatial abilities, whereas geographers had relatively high self ratings of environmental spatial abilities and engineers had relatively high self ratings of small-scale spatial abilities. Other scientific disciplines did not differ from the mean in self reported spatial abilities. Self ratings of verbal ability were uncorrelated with self ratings of spatial abilities and, as expected, were highest for humanities specialists.

Keywords: spatial ability, visualization, scale, science disciplines, STEM fields.

1 Introduction

Spatial thinking appears to be central to many scientific domains. For example, geologists reason about the physical processes that lead to the formation of structures such as mountains and canyons, chemists develop models of the structure of molecules, and zoologists map the tracks of animals to gain insights into their foraging behavior. Thus it is not surprising that measures of spatial ability are correlated with success in various scientific domains, such as physics [1, 2], chemistry [3], geology [4], mathematics [5], engineering [6], and medicine [7, 8]. A recent longitudinal study of gifted high school graduates indicated that spatial ability was an important predictor of their participation in science and engineering disciplines 15 years later, after controlling for SAT mathematics and verbal scores [9].

The measures of spatial ability that have been found to be predictive of success in science have been standardized paper-and-pencil tests measuring the ability to imagine transformations (mental rotation and paper folding). Recent research has found a

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 85–94, 2010.

partial dissociation between performance on these object-based (small-scale) spatial ability measures and individual differences in environmental spatial cognition tasks, such as learning the layout of a new environment, retracing a route, or pointing to unseen locations in a known environment [10]. For example, a review of 12 studies examining individual differences in a variety of tasks at the object and environmental scales of space found that the median correlation between these two types of spatial tasks was less than .3 and was not statistically significant in most of these studies [11]. Using structural equation modeling to control for task-specific and error variance, Hegarty, et al. [10] concluded that small-scale spatial abilities and the ability to learn the layout of an environment are partially dissociated; that is, they share some variance but are also dissociated. In their study the path coefficient linking large- and small-scale spatial abilities was .5, suggesting that they share 25 percent of variance.

The partial dissociation between small-scale and environmental spatial ability measures raises questions about whether scientists excel on both types of measures or only on small-scale spatial ability measures, as has been shown in previous research [1, 3, 4, 5, 8]. There are reasons to expect that there might be a dissociation between large-scale and small-scale spatial abilities across different scientific disciplines. Scientists in different disciplines study phenomena that occur at different scales of space [12]. For example, astronomers study large distances on the order of light years, often unobservable directly. Geologists study relations that span a range of scales: planet-wide tectonic plates, mountains, and even the microscopic spatial organization of mineral grains. In contrast, physiologists study phenomena at the scale of human and other organisms. Chemists examine reactions that occur at the miniscule scale of molecules, much too small to see, although they often use molecular models (manipulable objects) to represent these molecules. Might these sciences be differentially dependant on large- and small-scale spatial abilities?

The goals of this research project were to gather preliminary data with respect to the following research questions:

- Does scientific thinking depend equally on large- and small-scale spatial abilities?
- Do different scientific disciplines differ in their dependence on large- and small-scale spatial abilities?

To begin to answer this question, this paper presents preliminary evidence from an on-line questionnaire study in which scientists in different disciplines and specialists in other disciplines and professions completed self-report questionnaires regarding spatial abilities at different scales, and self-report measures of verbal ability. Although self-report measures are indirect measures of ability, they have been shown to have predictive validity. These measures range from those that simply ask people "how good is your sense of direction" to multi-item scales, such as the Santa Barbara Sense of Direction Scale, in which people report how much they agree with statements such as "I am very good at judging distances" or "I very easily get lost in a new city." Measures of self-reported sense of direction prove to be reliable and predict performance in such objective tasks such as ability to update one's position and orientation while walking blindfolded or in visually impoverished environments, to learn the layout of a new environments, and to judge the direction (in environment-based coordinates) from which a photograph of a familiar landmark had been taken [13, 14, 15].

Self-report measures are typically better predictors of performance in large-scale spatial cognition tasks than paper-and-pencil tests of spatial ability [10, 13].

We developed new self-report measures of small-scale spatial ability (the Philadelphia Spatial Abilities Scale) and verbal ability (the Philadelphia Verbal Abilities Scale) in the context of this research. These questionnaires, along with the Santa Barbara Sense of Direction Scale, were administered to scientists and individuals from other disciplines and professions in an on-line questionnaire. We predicted that scientists should rate themselves as higher on small-scale spatial abilities than individuals in non-scientific disciplines (humanities and other professions) reflecting previous evidence that performance in scientific disciplines is correlated with performance on small-scale (object-based) spatial abilities tests. We examined whether they also rated themselves higher on large-scale spatial abilities. Finally it was important to rule out the possibility that high self ratings on spatial abilities merely reflect general self esteem or high ratings of general ability. Thus we predicted that scientists would not rate themselves higher than average on verbal abilities and that self ratings of verbal ability would be dissociated from self ratings of spatial abilities.

2 Method

Participants. The participants were 850 individuals (485 female, 365 male) who responded to an email request to take an on-line survey. Four participants were currently in high school, 50 reported their level of education as high school, 237 had a bachelor's degree, 285 had a master's degree, and 274 had a doctorate. Given our interest in disciplinary specialization, the data from participants without a college degree ($N = 54$) were not analyzed further. The final sample that was analyzed had 796 participants (457 female, 339 male) with a mean age of 36.0 years ($SD = 13.0$).

Materials. The participants completed three questionnaires

(1) *Santa Barbara Sense of Direction Scale (SBSOD).* Participants were administered the Santa Barbara Sense of Direction Scale, which consists of 15 Likert-type items adapted from previous self-report scales of environmental spatial abilities [13]. Each item was a self-referential statement about some aspect of environmental spatial cognition; participants responded by clicking on a number from 1 ("strongly agree") to 7 ("strongly disagree") to indicate their level of agreement with each statement. The items are phrased such that approximately half of the items are stated positively, half negatively. An example of a positively stated item is "I am very good at judging distances"; an example of a negatively stated item is "I very easily get lost in a new city". In scoring, positively stated items were reversed so that a higher score indicates a better sense of direction. Sums of the 15 items were used for the analyses. The internal reliability (Cronbach's alpha) for this administration of the scale is .89.

(2) *The Philadelphia Spatial Abilities Scale*. The PSAS is a 25 item, Likert-style questionnaire that assesses reasoning in four categories of common spatial tasks [16]: static relations within objects (e.g., what the inside of an apple looks like), relations among objects (e.g., does a car fit into a parking spot), relations within deforming object (e.g., what a crushed can will look like) and relations among moving objects (e.g., putting together furniture) [17]. For this study, the PSAS was shortened to 16

questions. Each of the items asked the participants to assess their ability to complete a spatial task (e.g., "I am good at determining if my car fits into an available parallel parking spot") by selecting a number from 1 (strongly agree) to 7 (strongly disagree). The PSAS contains both negative statements (I have trouble with…) and positive statements (I am confident I can…), to avoid participants answering all items with "1". Questions spanned different types of spatial changes including rotation, crushing, and construction. While addressing various small-scale spatial skills, the instrument maintains a strong internal consistency, Cronbach's alpha = .87 for this administration of the 16-item scale.

The PSAS has good predictive validity for scores on paper-and-pencil tests of object transformation (Vandenberg and Kuse MRT) and disembedding (ETS hidden figures test). Overall score on the PSAS correlates .38 and .32 with the MRT and the hidden figures test, respectively [17].

(3) *The Philadelphia Verbal Abilities Scale*. The PVAS is a 10 item, Likert-style questionnaire that assessed verbal abilities, which was developed in the context of this research project. Each of these items asked the participants to assess a common verbal ability (e.g. "I am good at crossword puzzles" and "I would rather read a text explanation than look at a drawing or figure.") by selecting a number from 1 (strongly agree) to 7 (strongly disagree). The PVAS contains both negative statements, (I often have trouble…) and positive statements (I am good at…). Questions included examples of receptive (e.g., I would rather read), generative (e.g., I am good at expressing what I mean in words), and problem solving (e.g., I am good at Scrabble) skills. The instrument has a strong internal consistency reliability, Cronbach's alpha = .80 for this administration of the test.

Procedure. Participants were recruited by email, through mailing lists associated with the Spatial Intelligence and Learning Sciences (SILC) and Spatial@UCSB, by postings on the Web sites of the authors and their colleagues, and through personal contacts of the authors in different disciplines. People contacted by these methods were encouraged to forward the email request to their colleges, especially those in the physical sciences.

Invited participants visited a publicly available Web site, which presented each with brief instructions and the three-part survey form. First, participants were asked to provide their sex, age (optional), country, highest level of education completed, field of study, and subfield of study. The list of fields and subfields was a slightly restructured version of the list used by the National Opinion Research Center at the University of Chicago to conduct its annual Survey of Earned Doctorates.[1] In the next section, the 15 SBSOD items and the 16 PSAS items were combined together and randomly ordered for each administration of the survey. And in the final section, the 10 PVAS items were also randomly ordered.

After completing the survey, participants were presented with three dials plotting their mean sense of direction, small-scale spatial, and verbal skill ratings (on a scale of 1 to 7).

---

[1] http://www.norc.org/projects/Survey+of+Earned+Doctorates.htm

Coding. Scripts on the Web server recorded each administration of the survey to a database. Ratings were stored in their original form, but when calculating totals and means, ratings were reverse coded for negatively worded items. That is, on a question like "I am very good at giving directions," a rating of 1 ("strongly agree") was reverse scored to be 7. Thus in all cases, larger values on the totals, means, and resulting analyses indicate a self report of higher of ability.

3 Results

Descriptive Statistics. Means and standard deviations for the three scales are given in Table 1. The correlation between the two spatial ability scales was in the moderate to high range (r = .61, p <.001). Neither the Santa Barbara Sense of Direction Scale (r = .04, p = .30) nor the Philadelphia Spatial Abilities Scale (r = .00, p = .97) was correlated with the verbal abilities scale.

Table 1. Descriptive statistics for the three self-report scales

Scale	Possible Range	Mean	SD
Santa Barbara Sense of Direction (SBSOD)	15-105	74.36	16.66
Philadelphia Spatial Abilities (PSAS)	16-112	79.18	15.57
Philadelphia Verbal Abilities (PVAS)	10-70	48.99	9.47

Table 2 shows the number of respondents by field for respondents with a bachelor's, master's, and doctoral degree. Participants listing their field as physics, chemistry or astronomy were categorized as physical scientists. Biological sciences included those listing biomedical fields as their major area of specialty. The category of geoscientists included specialists in geology, oceanography, and meteorology. Finally, the category of professional fields included education, business, law, and health sciences.

Table 2.

	Bachelor's	Master's	Doctorate	Total
Physical Sciences	16	18	44	78
Biological/Biomedical	30	22	31	83
Geosciences	19	25	32	76
Geography	5	14	10	29
Engineering	22	23	14	59
Computer/Information Science	12	26	31	69
Psychology	35	45	51	131
Social Science	18	18	13	49
Humanities	26	41	16	83
Professional Fields	51	47	26	124

Differences between Specialties in Self-Reported Sense of Direction. To compare self-reported abilities across the different specialties, the scores on the three self-report scales were converted to standardized scores (z scores). Figure 1 shows the mean standardized scores by specialty for the Santa Barbara Sense of Direction Scale. The scores differed significantly by specialty, F (9, 771) = 3.12, p = .001. Geoscientists clearly had the highest self-ratings on this ability. Post-hoc (Tukey) tests indicated that geoscientists rated their sense of direction significantly higher than did biological scientists ($p < .001$), computer/information scientists ($p < .05$), humanists ($p < .01$), and those in professional fields ($p < .01$). No other differences between the professional groups were statistically significant and the pattern of results was very similar when only those with an advanced degree (master's or doctorate) were included in the analyses.

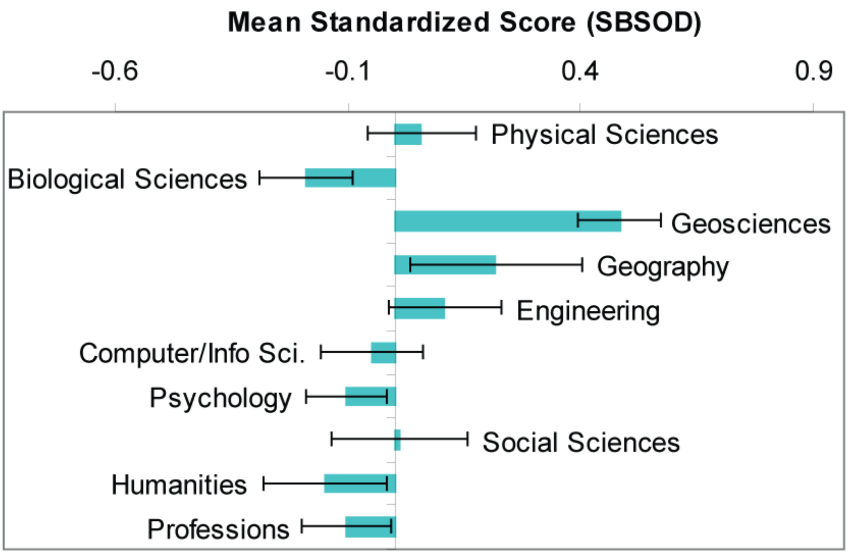


Fig. 1. Mean standardized score by specialty on the Santa Barbara Sense of Direction Scale by discipline for all participants with a college degree

Differences between Specialties in Self-Reported Small-Scale Spatial Ability. Figure 2 shows the mean standardized scores by specialty for the Philadelphia Spatial Abilities Scale. The scores differed significantly by specialty, F (9, 771) = 4.95, $p < .001$. Again geoscientists had the highest self-ratings on this ability. Compared to sense of direction, engineers rated their object-based spatial abilities as somewhat higher, while geographers rated their object-based spatial abilities as somewhat lower. Post-hoc (Tukey) tests indicated that geoscientists rated their object-based abilities as significantly higher than did biological scientists ($p < .01$), computer/information scientists ($p < .05$), psychologists ($p < .001$), social scientists ($p < .01$), humanists ($p < .01$) and those in professional fields ($p < .001$). Engineers also rated their object-based spatial abilities as significantly higher than those of psychologists ($p < .05$) and those in professional fields ($p < .05$). No other differences were statistically significant and

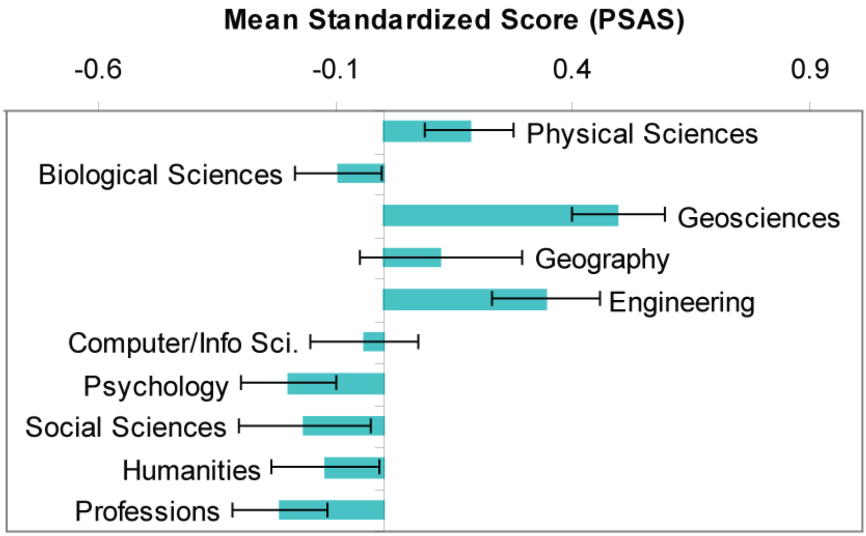


Fig. 2. Mean standardized score by specialty on the Philadelphia Spatial Abilities Scale by discipline for all participants with a college degree

the pattern of data was very similar when we considered only those participants with an advanced degree (master's or doctorate).

Differences between Specialties in Self-Reported Verbal Ability. Finally, Figure 3 shows the mean standardized scores by specialty for the Philadelphia Verbal Abilities

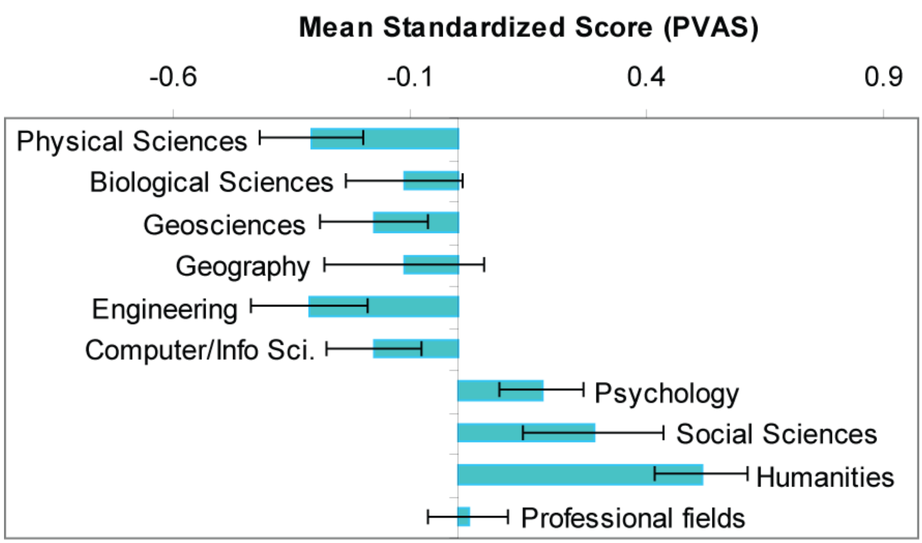


Fig. 3. Mean standardized score by specialty on the Philadelphia Verbal Abilities Scale by discipline for all participants with a college degree

Scale. It can be seen that the pattern is very different from those observed for the spatial ability scales, indicating that participants discriminated between the abilities and did not merely rank themselves as high or low on all abilities. Furthermore the relative ratings were as expected, with humanities specialists rating themselves highest on verbal ability. The ratings differed significantly by specialty, $F(9, 771) = 5.74$, $p < .001$. Post-hoc (Tukey) tests indicated that humanists rated their verbal abilities as significantly higher than did biological scientists ($p < .01$), computer/information scientists ($p < .001$), geoscientists ($p < .001$), engineers ($p < .001$), and physical scientists ($p < .001$). Social scientists and psychologists also rated their verbal abilities higher than did physical scientists and engineers ($p < .05$). No other differences were statistically significant. When we considered only those with an advanced degree, participants in professional fields rated their verbal abilities as higher (mean standardized score = .25) but otherwise the patterns were very similar.

4 Discussion

In summary, this study provides evidence that people's self assessments of their environmental (large-scale) spatial abilities are not completely parallel to their self assessments of smaller scale object-based spatial abilities. This pattern of partial dissociation between spatial abilities at small and large scales—found here using self-report questionnaires—is similar to that found when measuring people's objective performance [10]. Second, it shows that all sciences are not equal in terms of self-reported spatial abilities. While geoscientists have the highest ratings on both large-scale and small-scale spatial abilities, geographers have the second highest ratings on measures of large-scale spatial abilities and engineers and physical scientists rate themselves as higher than geographers on small-scale spatial abilities. While previous research has highlighted the importance of spatial ability in various scientific disciplines, this study raises the question of whether spatial ability is equally important for success in all sciences.

Our results support the validity of the self-report scales. They provide discriminant validity in that the mean self-report ratings for the two spatial abilities scales are uncorrelated with the self-report ratings for the verbal abilities scale. Thus, participants were not merely rating themselves as high or low on all abilities. The fact that the two spatial ability scales were somewhat correlated provides convergent validity, as we would expect these two scales to be more highly correlated with each than with verbal ability. At the same time, it was clear that participants differentiated small-scale from large-scale spatial abilities in that the correlation of the two spatial ability scales was in the moderate (not high) range. Finally, there was evidence for criterion validity of the scales in that the scientists who ranked themselves as high in environmental spatial ability were from disciplines that deal with larger-scale spaces, while humanities specialists rated themselves as highest on verbal abilities.

It is perhaps surprising that physical scientists (physicists, chemists and astronomers) do not rank themselves more highly in spatial abilities, given that spatial ability is often assumed to be a prerequisite for success in physical sciences. This group was made up of 54 physicists, 19 chemists and 5 astronomers, and a further breakdown into these 3 groups did not reveal any significant differences between these sciences, although this analysis is limited by low power. However, recent research suggests that the development of expertise in science is often accompanied by the acquisition of analytic heuristics for solving spatial problems. For example, Stieff [16] found that beginning organic chemistry students almost always used a mental rotation strategy when determining whether two molecules had the same structure, but expert chemists were much more likely to use an analytical strategy, especially when the molecule was symmetrical. A post-hoc analysis revealed no differences in self-rated abilities of physical scientists as a function of education level, but again this analysis is limited by low power. In examining the use of visualization versus analytical strategies in domains such as mechanics and chemistry, researchers have suggested that visual-spatial strategies are default domain-general problem solving heuristics that are used by novices or by experts in novel situations, whereas rule-based analytic strategies are learned or discovered in the course of instruction and are used by experts in routine problem solving [18, 19].

Our research suggests that geoscientists (geologists, oceanographers, and meteorologists) appear to be most dependent on spatial abilities in that geoscientists had the highest self ratings on both large-scale and small-scale spatial abilities. These disciplines may be particularly dependent on spatial abilities because they are more grounded in real-world experiences of spatial structures and processes, such as rock configurations, ocean waves, and thunderstorms. Geologists must keep track of spatial locations at the environmental scale when they go on field trips and observe outcrops to reason about the spatial processes that give rise to structures and processes at the earth's surface. They also depend more on spatial representations than do other scientists, in that their publications have a greater ratio of graphics to text area than other sciences [20]. While chemists, for example, are also highly dependent on graphical representations and have invented many different diagrammatic conventions for molecules, the structures and spatial processes that they study are less tied to real-world spatial experience, in that they do not have direct experience of the space of molecules.

While intriguing, the results of this study have several limitations. First, they are based on self reports of ability, which at best are only a reflection of objective differences in spatial intelligence. It will be important to follow up these results with studies that measure the abilities of scientists and other specialists more objectively. Second, our results are based on responses from a convenience sample rather than a truly representative sample of the populations being compared. Finally, the data are correlational, so they cannot be interpreted as evidence for a causal relationship between levels of spatial abilities and participation in sciences or vice versa. Nevertheless, our results are important in highlighting the need for studies of the role of spatial thinking in science to be mindful of both varieties of spatial intelligence and varieties in the spatial demands of different sciences.

Acknowledgements. This research was funded in part from National Science Foundation grant 0722333 to Mary Hegarty, National Science Foundation Grant 0823557 to Thomas F. Shipley, and National Science Foundation Science of Learning Center Grant 0541957 that supports the Spatial Intelligence and Learning Center (SILC).

References

1. Kozhevnikov, M., Motes, M., Hegarty, M.: Spatial visualization in physics problem solving. Cognitive Science 31, 549–579 (2007)
2. Sorby, S.A.: A course in spatial visualization and its impact on the retention of female engineering students. Journal of Women and Minorities in Science and Engineering 7, 153–172 (2001)
3. Coleman, S.L., Gotch, A.J.: Spatial perception skills of chemistry students. Journal of Chemical Education 75, 206–209 (1998)
4. Orion, N., Ben-Chaim, D., Kali, Y.: Relationship between earth-science education and spatial visualization. Journal of Geoscience Education 45, 129–132 (1997)
5. Casey, M.B., Nuttall, R.L., Pezaris, E.: Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. Developmental Psychology 33, 669–680 (1997)
6. Peters, M., Chisholm, P., Laeng, B.: Spatial ability, student gender and academic performance. Journal of Engineering Education 84, 69–73 (1995)
7. Schueneman, A.L., Pickelman, J., Hesslein, R., Freeark, R.J.: Neuropsychological predictors of operative skill among general surgery residents. Surgery 96, 288–295 (1984)
8. Keehner, M.M., Tendick, F., Meng, M.V., Anwar, H.P., Hegarty, M., Stoller, M.L., Duh, Q.: Spatial ability, experience, and skill in laparoscopic surgery. The American Journal of Surgery 188, 71–75 (2004)
9. Shea, D.L., Lubinski, D., Benbow, C.P.: Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. Journal of Educational Psychology 93, 604–614 (2001)
10. Hegarty, M., Montello, D.R., Richardson, A.E., Ishikawa, T., Lovelace, K.: Spatial Abilities at Different Scales: Individual Differences in Aptitude-Test Performance and Spatial-Layout Learning. Intelligence 34, 151–176 (2006)
11. Hegarty, M., Waller, D.: Individual differences in spatial abilities. In: Shah, P., Miyake, A. (eds.) Handbook of Visuospatial Thinking, pp. 121–169. Cambridge University Press, Cambridge (2006)
12. Montello, D.R.: Scale and multiple psychologies of space. In: Frank, A.U., Campari, I. (eds.) COSIT 1993. LNCS, vol. 716, pp. 312–321. Springer, Heidelberg (1993)
13. Hegarty, M., Richardson, A.E., Montello, D.R., Lovelace, K., Subbiah, I.: Development of a self-report measure of environmental spatial ability. Intelligence 30, 425–447 (2002)
14. Kozlowski, L.T., Bryant, K.J.: Sense-of-direction, spatial orientation, and cognitive maps. Journal of Experimental Psychology: Human Perception and Performance 3, 590–598 (1977)
15. Sholl, M.J., Kenny, R.J., Dellaporta, K.: Allocentric-heading recall and its relation to self-reported sense-of-direction. Journal of Experimental Psychology: Learning, Memory & Cognition 32, 516–533 (2006)
16. Chatterjee, A.: The Neural Organization of Spatial Thought and Language. Seminars in Speech and Language 29, 226–238 (2008)
17. Crookes, R.D., Shipley, T.F., Hegarty, M.: Development of a self-report measure of spatial reasoning abilities. Poster presented at the annual meeting of the Eastern Psychological Association Meeting, Brooklyn, NY (March 2010)
18. Stieff, M.: Mental rotation and diagrammatic reasoning in science. Learning and Instruction 17, 219–234 (2007)
19. Schwartz, D.L., Black, J.B.: Shuttling between depictive models and abstract rules: Induction and fall-back. Cognitive Science 20, 457–497 (1996)
20. Kastens, K.: Synthesis of research on thinking & learning in the geosciences: Developing representational competence. Presented at the annual meeting of the Geological Society of America, October 18-21 (2009)

Gestures in Geology: The Roles of Spatial Skills, Expertise, and Communicative Context

Lynn S. Liben[1], Adam E. Christensen[1], and Kim A. Kastens[2]

[1] Department of Psychology,
The Pennsylvania State University, University Park, PA 16802, USA
[2] Lamont-Doherty Earth Observatory, Department of Earth &
Environmental Sciences and Columbia University, Palisades, NY 10964 USA
liben@psu.edu, aec187@psu.edu, kastens@ldeo.columbia.edu

Abstract. Spatial thinking is an essential part of science, technology, engineering, and mathematics (STEM), domains that entail external spatial representations such as 2D graphics, 3D models, and—the focus here—gestures. University students (a) read aloud information about the geological concepts of strike and dip, (b) completed strike and dip tasks, and (c) explained these concepts to another student via audio and video. Gestural patterns varied across reading, audio, and video contexts in interaction with participant variables of spatial skill and prior geology coursework. Only novices gestured during reading, interpreted as novices' attempts to aid their own conceptual understanding of new information. All participant groups produced different gestural patterns in audio versus video contexts, suggesting the communicative intent of many gestures.

Keywords: Spatial cognition, gesture, spatial representation, geology, science education.

1 Introduction

The role of spatial thinking in science, technology, engineering, and math (STEM) fields has been attracting increased attention in both cognitive science and education [1]. In part, spatial thinking involves building, manipulating, and using *mental* spatial representations as when a chemist—trying to determine whether two molecules are the same or different—creates a mental image of one molecule and then mentally rotates and superimposes it on an image of a second molecule [2] or as when a geologist—trying to develop a hypothesis about the formation of the physical landscape—imagines how a cross-section of a rock outcrop would look. Spatial thinking also involves creating and using *external* spatial representations [1, 3]. These include two-dimensional (2D) graphic representations such as maps [4], three-dimensional (3D) models such as stream tables in Earth Science classes [5], and iconic gestures, the focus of the work described here. Iconic gestures are those in which hand or body positions mimic referents' configurations or movements. Well-documented examples include dynamic hand gestures representing the Earth's movement and orientation during seasonal change [6] or movements of gears in a pulley system [7].

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 95–111, 2010.

The STEM domain in which we have been conducting our work is geology, a highly spatial science that makes heavy use of graphics [8, 9], models [10], and gestural representations [11, 12]. One major purpose of our program of research has been to study the creation and use of external representations as people attempt to learn or enact scientific skills or concepts. Of interest is identifying what task conditions differentially encourage the creation and use of external representations, what characteristics of individuals (e.g., different levels of spatial abilities, different domain-relevant expertise) are linked to differential use of external representations, and how representations found in educational materials are used and interpreted.

Within geology, we have been concerned primarily with how individuals observe, record, and reason about spatial orientations of rock outcrops. Some of our work has been aimed at studying how individuals approach complex, higher-level inferential reasoning tasks. Illustrative is research in which participants were taken to a realistically-scaled field area in which eight artificial (plywood) layered outcrops had been installed and, after visiting all eight outcrops, were asked to infer what geological structure might lie beneath the outcrops [13]. To succeed on this task, participants must observe and represent (mentally or graphically) the orientations and inclines of the individual outcrops, and then systematically and logically integrate the individual observations to draw a spatial inference. This research has shown that participants use a variety of environmental representations (sentential notes, sketch maps, 3D models, and gestures) as they observe and reason with geological information [12].

Some of our work has been designed to study specific component processes entailed within higher-level inferential tasks [14]. In particular, we have investigated how individuals observe and record a layered rock outcrop's *strike*—its orientation with respect to compass direction—and *dip*—its inclination with respect to the horizontal plane. Findings from prior work on the component concepts of strike and dip [14] demonstrated that after being given introductory material about these concepts, many college students made dramatic errors in recording an outcrop's strike and dip on a campus map. Errors were especially prevalent among students who had low performance on the water level task (WLT), a task that taps the respondent's ability to represent the invariant horizontality of water within a tilted container [15].

The research described here was designed to refine and extend our prior work by studying the use of a particular kind of external representational medium—gesture—as individuals tried to understand and communicate about the component concepts of strike and dip. More specifically, one purpose of our research was to study the incidence and types of gestures produced under three different task contexts—one focused on participants' own understanding of strike and dip, and the other two focused on explaining strike and dip to another person via either audio or video recordings. A second purpose of this research was to examine possible differences in gesturing in relation to two participant variables—performance on the WLT and pre-experimental instruction on strike and dip. The former was motivated by the finding, mentioned earlier, that students performed differently on the geology tasks in relation to their performance on the WLT [14]. Here we asked whether different water-level performance would also be associated with different gestural profiles. With respect to the latter, we collected data from students who had no prior instruction in the concepts of strike and dip (referred to as "novices") as well as from students who had some prior instruction in these concepts from coursework in geology classes (referred to as "geo-students").

All participants were first observed while they read aloud a lesson on strike and dip. Of interest was whether participants would translate the verbal and graphic descriptions included in the lesson into gestural representations as a way to support their own comprehension. We anticipated that self-directed gestures accompanying reading would be particularly likely to occur among those who were not already familiar with the geological concepts of strike and dip, that is, among the novices.

After reading the lesson, participants were asked to complete a series of model (table-top) strike and dip tasks. These tasks were included to provide an objective assessment of how well participants had understood the instructional lesson. Next, participants were asked to explain strike and dip to another person by producing audio and video recordings. We reasoned that to the degree that participants' gestures during these communication tasks were truly other- rather than self- directed, we would observe a higher incidence (and perhaps different kinds) of gestures in the video than in the audio context. To the degree that gestures during these tasks were being used to support participants' own understanding, gestures would be expected to be comparable across both communicative contexts.

Given evidence in the gesture literature that those with higher spatial skills tend to gesture more [16], we expected to find greater use of communicative gesture among those with higher WLT scores. We did not, however, have strong directional predictions about relative amounts of gesturing by geo-students versus novices given that there was reason to expect considerable gesturing in both groups: educational research has shown that experts routinely make heavy use of representational gestures [17, 18, 19] while cognitive-developmental research has shown significant gesturing in children on the cusp of understanding new concepts [20, 21, 22].

2 Method

2.1 Participants

Participants classified as novices were students recruited from the psychology subject pool of a large, public university in the United States who self-reported no prior experience with the geological concepts of strike and dip. Complete data were collected from 40 men and 37 women (*M* [*SD*] age = 19.4 [1.4] years) in this novice group. Three additional novices had been recruited but did not contribute data, in one case because the participant did not wish to be videotaped and in two cases because of a failure to understand the initial directions about reading the geology lesson. Participants classified as geo-students included 5 students from the same psychology subject pool who self-reported having learned about strike and dip in a geology class, plus 11 undergraduate and 5 graduate students who responded to various recruitment efforts (e.g., researcher visits to geology classes, e-mails to all geoscience majors). The final geo-student sample included 14 men and 7 women, *M* [*SD*] age = 22.7 [3.7] years. Participants recruited from the subject pool received extra course credit for participation; those recruited in other ways received $12. Mirroring the demographics of the university, almost all participants were white.

2.2 Procedural Overview

Participants were greeted in a small room (3.0 m X 3.6 m) in the psychology department where they first completed consent procedures that included giving permission to have the entire session videotaped. Participants then completed the WLT followed by four activities related to strike and dip: (1) reading instructions about strike and dip, (2) completing tabletop model strike and dip tasks, (3) recording either an audio or video explanation of strike and dip for another person, and (4) recording a second explanation in the format not already used (i.e., audio then video or the reverse).

2.3 Water-Level Task

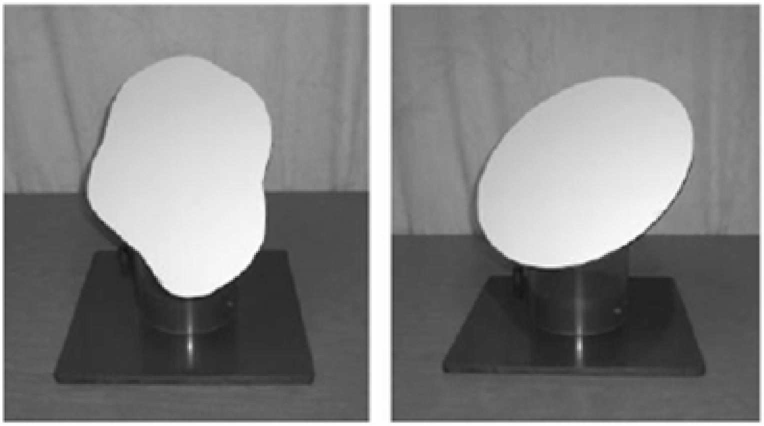
For the WLT, participants were given drawings of six straight-sided bottles tipped from upright by 30°, 45°, and 60° to the right and left [23]. Students were asked to draw a line inside each bottle to show "where the water would be if the bottle were about half full and held in the position shown." Lines were measured for degrees deviation from horizontal, and those drawn more than 5° off horizontal were categorized as errors. Based on earlier research [24, 25, 26], participants with zero or one error were classified as falling into the high water-level group (WLG) and those with two or more errors were classified as falling into the low WLG. As is routinely true in research using this task that examines performance in relation to participant sex [24, 27, 28, 29, 30], proportions of males and females falling into the two groups differed: 74% of women but only 39% of men were in the low WLG. As would also be expected from research showing that students in more spatially rich and demanding disciplines perform better on an array of spatial tasks including the WLT [4, 27, 31, 32], proportions of novices and geo-students falling into the two groups differed as well: 61% of the novices but only 38% of the geo-students were in the low WLG.

2.4 Introduction to Strike and Dip

After completing the WLT, participants were told that they were going to be asked to learn about geological concepts called strike and dip. They were asked if they had any geology experience and if so, to specify their experience. They were then given written instructions about these geological concepts that had been adapted from the United States Geologic Survey [33] website for use in earlier research [14]. As explained to participants, strike refers to the line formed by the intersection of a horizontal plane and a slanting rock surface. The strike line recorded on a geological map represents the orientation of the slanted surface, conventionally expressed as degrees clockwise from north (i.e., its azimuth). Thus, for example, an east-west trending line drawn on the map would indicate that the outcrop is oriented in an east-west direction. A line drawn on the map 30° clockwise from due north would indicate that the outcrop is oriented in a NNE-SSW direction.

The dip line is drawn on the map perpendicular to the strike line to show the direction in which the rock surface is sloping. Thus, for the east-west strike example, if the rock surface sloped downward toward north, the dip line would be drawn on the north side of the strike line. A number marked next to the dip line is used to indicate the steepness of the slope in degrees, measured downward from the horizontal. Verbatim wording and figures used in the explanation are provided elsewhere [14].

To allow us to observe when participants were examining images during this phase of the study, figures from the instructions were posted on an easel at one end of the table (rather than appearing within the instruction booklet as they had in earlier research). At the other end of the table was the model outcrop shown in Figure 1a.



(a) (b)

Fig. 1. (a) Model outcrop available during reading and explanation (audio; video) phases and used for strike tasks. (b) One of the five models used for the dip tasks. During tasks, fabric was fitted around the round surface and draped over the models' bases (see text).

To explain the motivation for the research and to establish a context in which we could make observations about students' processing of the instructional material, participants were told:

> *For this first part we are going to have you learn about some concepts in geology called strike and dip. In the past, we have introduced people to these concepts by reading material to them aloud, but some people have found the explanations hard to understand. We may have been reading too fast or not emphasizing the right parts. As one way to learn more about pacing and what to emphasize, we are now asking you to read the information aloud to yourself. You should just read it at whatever pace seems best for you to understand the material. Also, you should feel free to back up and re-read any sections that you want. We basically want you to read the instructions to yourself the way you might read a text passage that you are trying to understand. We're only asking you to do the reading aloud so that we will be able to see where people take more time, pause, or re-read things.*

The experimenter mentioned that the pictures referred to in the instructions were on the poster "so that you can consult these as you go or return to them as needed" and pointed out that there was a model of a rock outcrop on the table in case you "find it helpful." The experimenter foreshadowed later tasks by saying "Keep in mind also that you will actually be completing some of the tasks you are learning about on models in this room and you will have to try to explain it to someone else as well." The experimenter left the room, telling participants to take as much time as needed, and to let the experimenter know when they had finished. The experimenter then left the room and waited in the hallway while the participant read the instructions. Of interest here was the use of gesture during reading; elsewhere we will examine data concerning reading speed, pause patterns, repetitions, speech disfluencies, the relation between verbalizations and gestures, and other indices of processing the materials.

2.5 Strike and Dip Tasks

After participants informed the experimenter that they had read the instructions and the experimenter had returned to the testing room, all materials were cleared from the table. Participants were told that they would be doing some geology tasks, but that because they were inside they would be doing them on models rather than on real rocks. All participants were given the dip task first because it did not require introducing a map and because earlier research [14] had shown no effect of task order.

For the dip task, participants were shown a round model like that shown in Figure 1b. For all dip items, fabric was attached to the edges of the surface to mimic the way a rock surface would emerge from the soil and to obscure the immediate orthogonal cues of the model's base. Participants were told that they would be viewing five surfaces like the one in front of them, and that for each item they should estimate the dip angle. They were given a picture of a protractor simply "to remind you of angle sizes" but asked to refrain from trying to use it "to measure any of the angles directly." Participants were also given a set of multi-page instructions that contained the text read earlier into which the figures from the poster had been re-inserted. Participants were explicitly told that "you can refer back to these as much as you need."

For each dip item (inclines at 15°, 30°, 45°, 60°, and 75°), the model was placed on the table so that from the participant's position, the full frontal view of the planar surface was rotated 30° counterclockwise. The five models were presented in random order for each participant. As participants worked on the task, the experimenter used a check list to record observable strategies (e.g., use of an object such as the clipboard to estimate a plane, holding the palm flat on the surface of the model). After giving verbal estimates of the dip angle for all five items, participants provided confidence ratings using a scale from 1 ("completely uncertain") to 7 ("completely confident") and were then asked open-ended questions about how they had accomplished the task. Given space constraints, neither strategy nor confidence data are discussed here.

For the strike task, the amorphous-shaped model shown in Figure 1a was placed at particular locations and orientations on the table with a cloth draped around the base to obscure proximal orthogonal cues. Participants were told:

> *For these next tasks you're going to be drawing the strike lines as described in the instructions. Again, I want you to imagine this angled surface is the surface of rock. I'm going to place the surface in 8 different orientations. For each one I want you to mark the strike line for that orientation on a map I will give you.*

Participants were then given a clipboard with a scaled (15:1), black-and-white plan room map that depicted walls, a desk, table, bookshelf, row of filing cabinets, and doorway. The clipboard was handed to participants so that the map was aligned with the space. In addition, the experimenter oriented the participant to the map by pointing out the correspondences between the door and file cabinets in the room and the door and file cabinet symbols on the map. The rectangular table at which the participant worked was omitted from the map to avoid providing proximal table-frame cues for the strike lines. Although not told the reason for its omission, participants were alerted to the fact that the table itself was not represented on the map.

After the participant was satisfied with a given strike line, the map was removed from the clipboard to reveal a fresh map to be used for the next strike item. The model was then moved to the next location and placed in its designated orientation, and the participant was asked to draw the strike line on the new map. Orders in which the eight items were given were randomized for each participant. As was the case for the dip task, the experimenter used a check list to record observable strategies (e.g., noting whether the map remained aligned and if the participant looked back at the instructions), asked participants for confidence ratings, and queried participants about the way they had solved the task, although again, these data are not discussed here.

2.6 Explanation of Strike and Dip to Another Person

The final components of the strike and dip tasks involved asking participants to explain the geological concepts to another person. Specifically, participants were told:

> *In the next part of this study we are trying to see whether it would be easier for people to learn about strike and dip from another person rather than from all written instructions. So in this next part I'm going to ask you to explain the concepts to another student who hasn't taken geology.*

For the audio-only context, participants were told that they would next make an audio recording to explain to someone else how to determine strike and dip. Participants were told that only the audio track of the ongoing video recording would be used so that the other person "will not be able to see you or anything in the room—they will hear only your voice." Participants were told that the other person would not have the instruction book, but would have the same pictures as well as the same model.

With the expectation that at least some participants would feel that their own understanding was inadequate to teach the concepts to someone else, they were told:

> *Please explain strike and dip as well as you can. Explain whatever parts you understand and if you're unsure or can't explain anymore it's fine to say so. You can just say 'that's all I understand.' You are free to refer back to the instructions, but we do not want you to read them. We want you to try to explain these ideas in your own words. Maybe you have a different way of thinking about the problem, or anything you think will help the other person be able to understand strike and dip. I'm leaving the model on the table in case you find it helpful.*

Parallel instructions were used for the videotaped instructions except for specifying the video format and thus that the other student would be able to see them and the room. Half the participants were asked to prepare audio then video explanations; half followed the reverse order. In neither case were students told beforehand that they would be preparing a second version. As in the initial reading of the instructions, the experimenter stepped outside the room during this portion of the task, and participants were asked to let the experimenter know when they were finished.

2.7 Gesture Coding

Two major types of gestures were coded from videotapes—*deictic* and *iconic*. Deictic gestures are gestures judged to serve an attentional function, that is, gestures judged as directing someone's attention to something in the environment. Illustrative of deictic gestures in the current study were those in which the participant pointed to one of the photographs posted on the easel. Iconic gestures are gestures that represent spatial information of the referent via hand or body positions and movements. Among the specific types of iconic gestures coded in the current study were gestures in which the hands were used to represent 3D planar surfaces, as in making horizontal, vertical, or inclined planes with the hand in either a static or dynamic fashion, or as in shifting the hand repeatedly between horizontal and vertical positions (see Figure 2).

(a) (b) (c)

Fig. 2. Illustrative iconic gestures at each locus. (a) Participant made picture-locus gesture (at the image of the protractor) while saying, "About like that," referring to a 55° dip. (b) Participant made model-locus gesture while saying, "Dip is the degree from zero, so find the level part." (c) Participant made space-locus gesture of moving hands back and forth over one another while saying, "In other places the rocks are tilted and not in flat layers, like in Photo B."

In addition to coding the type of gesture, we also coded the locus of each gesture. Specifically, each gesture was coded with respect to whether it in some way incorporated or related to (a) a picture available on the poster (e.g., holding a flattened, vertically oriented hand up to the picture of the protractor to indicate a dip of 90°), (b) the outcrop model that was available on the table (e.g., holding a hand horizontally against the model surface), or (c) if it instead was made independently of the available representations and thus in the open space in front of the participant (e.g., holding the hands over one another to represent rock layers, separate from a picture or model). We refer to these three, respectively, as picture-, model-, and space-loci. Ionic gestures at each locus are shown in Figure 2. Note that gestures coded as iconic—even those with a picture or model locus—did not simply involve pointing to or touching the representation. To be coded as iconic, the gesture had to in some way provide some spatial information with a hand or body position or movement. If a participant merely pointed to or touched a picture or the outcrop model, the gesture would have been coded as deictic.

Interrater reliability of coding was assessed using Generalized Sequential Querier (GSEQ) [34]. With two independent coders scoring 52 of the 98 videos (53%), the overall kappa for the iconic codes was .81. When iconic gestures were broken down by locus, reliabilities for pictures, model, and space were .57, .88, and .73 respectively. For deictic gestures, overall kappa was .84; kappas for pictures, model, and space were .83, .74, and .63 respectively. All kappa values, with one exception, were thus in either the "substantial" (.61–.80) or "almost perfect" (.81–1.0) ranges according to the benchmarks suggested by Landis and Koch [35]. Because the current paper is focused on representational gestures, only findings from iconic gestures are reported here.

3 Results

The data are presented below in three sections. First, we describe participants' use of gestures while reading information about strike and dip. Second, we present data on participants' gestures during the communication task, discussing gestural patterns in relation to context (audio versus video), and in relation to participant variables (performance on the WLT and geology-relevant coursework). Third, we report the relation between participants' gestures and their performance on the model geology tasks.

3.1 Gestures while Reading Instructions Aloud

One purpose of this research was to examine the possibility that participants would produce gestures while reading the material on strike and dip, and if so, to learn whether the incidence and loci of gestures would differ in relation to two participant qualities—first, domain-specific experience as indexed by prior exposure to strike and dip in geology coursework, and second, level of spatial skills as assessed by the WLT.

Gestures in the self-reading portion of the investigation were rare: only 12 of the 98 participants made iconic gestures. All 12 were novices, meaning that 16% (12 of 77) of the novices generated iconic gestures during reading compared to 0% (0 of 21) of the geo-students, χ^2 (1, N = 98) = 3.73, Fischer exact $p = .045$. Of the 12 who gestured, 5 were women (4 in the low WLG) and 7 were men (1 in the low WLG). Some of those participants who gestured did so more than once. One particularly active participant, for example, gestured repeatedly: While reading about the strike line, she made a plane with her hand and lined it up with the strike line on the figure; while reading about the dip line, she made an inclined plane with her and then rotated it down to a horizontal position; while reading descriptions of dip angles of 90º and 0º, she held her hand flat, first vertically and then horizontally (repeating these gestures during two readings of the same text); while reading about a 60° dip, she made an inclined plane with her hand; and, finally, while reading about the protractor, she pointed one finger of one hand straight up in the air hand and used a finger on her other hand to point to the drawing of the protractor and to trace a line through the air, downward at an angle. Although a few participants thus made multiple gestures, gestures during reading were relatively rare and, as mentioned above, were confined to the novices. Thus, we could not conduct additional formal analyses of the incidence of gestures in relation to participant variables.

3.2 Gestures while Explaining Strike and Dip to Another Person

Gestures during the communicative phases of the study were also varied. Participants produced the full range of gesture types described earlier.

Overall, far more participants gestured during the communicative phases of the study than during the reading phase. Among the novices, 70% (54 of 77) gestured during the audio context and 92% (71 of 77) gestured during the video context. Among the geo-students, 67% (14 of 21) gestured during the audio context and 100% (21 of 21) gestured during the video context. Furthermore, participants routinely produced multiple gestures. As a consequence, the corpus of observed gestures was adequate to allow us to examine the use of representational gestures under differing communicative contexts and among participants with different spatial skills and geology experience. We thus conducted an analysis of variance (ANOVA) in which the numbers of iconic gestures served as the dependent measure. The between-subject factors were performance on the water level task (low vs. high WLGs) and prior experience with strike and dip (novices vs. geo-students). The within-subject factors were communicative context (audio vs. video) and locus of gesture (pictures, model, or space).[1] As necessary, t-tests were used for post hoc analyses, with alpha set at .05.

The ANOVA revealed that, overall, there was more gesture in the video than audio contexts and more gesture was produced by participants with higher water-level scores and with greater geology experience. These significant main effects are not discussed further, however, because they were subsumed under a four-way interaction, $F(2,188) = 4.63$, $p = .011$, shown graphically in Figure 3.

One particularly striking aspect of the interaction apparent in Figure 3 is the difference between novices and geo-students with respect to WLG. Among the novices, the patterns between audio and video are similar for both low and high WLGs. More specifically, within the novices, for both low and high WLGs, whatever iconic gesturing occurred generally took place in open space rather than at a graphic representation (a picture) or a 3D representation (the outcrop model). Novices did produce significantly more gestures at the model in the video than audio context, but even so, these remained significantly less common than gestures made in open space.

[1] Prior to conducting the analyses reported in the text, a preliminary analysis was conducted to test for possible order effects (audio then video vs. video then audio). There were neither main effects nor interactions involving order, and thus data were pooled across order for subsequent analyses. A second additional analysis was conducted to examine possible gender effects. There is a well-established link between gender and water-level performance that was, as reported earlier, evident in the current sample as well. However, because spatial skills (here assessed by the water-level task) are more directly relevant for this research than are other gender-linked qualities, and because some cells would have been unacceptably small if we were to have sub-divided participants by both gender and WLG, we collapsed over gender in the analyses reported in the text. We did, however, conduct an additional analysis on data from the novice participants because this sample was large enough to accommodate both gender and WLG as between-subjects factors. This ANOVA revealed significant effects of WLG but not gender, supporting the decision to focus on WLG in the main analyses reported in the text.

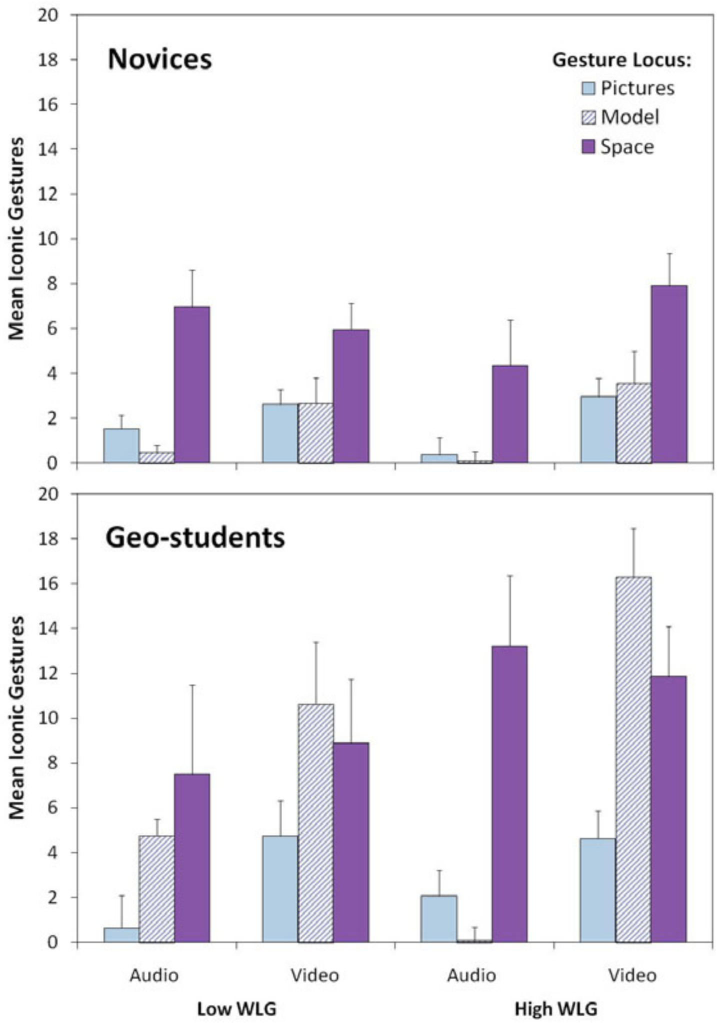


Fig. 3. Iconic gestures by student group, water-level group (WLG), context, and locus. Means are observed and error bars are model standard errors.

In contrast, within the group of geo-students, the gesturing pattern across the three loci differed dramatically by context in relation to participants' membership in the two WLGs. Especially striking is the high incidence of model-locus gestures, particularly in the video context. Interestingly, although both low and high WLG geo-students gestured at the model in the video context, only the high WLG geo-students avoided gesturing at the model in the audio context. Strikingly, within these geo-students, there was no parallel contrast in space-locus gestures in relation to context. That is, among the geo-students, the frequency of space-locus gestures was high in both audio and video contexts irrespective of geo-students' WLG membership.

3.3 Gestures in Relation to Performance on the Geology Tasks

The final set of analyses concerned the relation between participants' use of iconic gesture and their performance on the geological strike and dip tasks.

First, we analyzed data relevant to the possibility that participants' use of gesturing while reading the instructions would predict their performance on the geology tasks. Given how few gestures were used during the instruction-reading phase, we explored this association by comparing performance on model strike and dip tasks (using average error in degrees as the dependent measures) between those novices who had gestured versus those who had not gestured during the initial instruction-reading phase. Although mean errors were lower for gesturers than for non-gesturers on both strike (33.1° vs. 40.8°) and dip (12.0° vs. 15.2°), in neither case were these differences significant, $t(75) = 1.43$, $p = .156$ and $t(75) = 0.76$, $p = .448$, respectively.

Second, we explored data relevant to the possibility that better understanding of strike and dip would be associated with greater use of gesture during the communication tasks. Given the extensive use of gesturing in the communication phase (in contrast to its far more limited use during the instruction-reading phase), we were able to use gesture frequency (rather than a dichotomous grouping variable) as well as additional covariates in our analyses. Thus, we conducted two regression analyses to test whether performance on the geology tasks would predict iconic gesturing in the two communication tasks (audio and video) above and beyond whatever variance could be explained by the other participant variables expected to vary with gesture (spatial skill and prior geological coursework). For both analyses, on level 1 we entered water-level performance (mean degrees-error), on level 2 we entered domain experience (0 for novice and 1 for those with prior geology coursework), and on level 3, we entered strike and dip performance (mean degrees-error for each, separately). For the analysis in which the number of iconic gestures in the audio context served as the criterion variable, the only significant predictor was geological experience, $\beta = .23$, $p = .026$. For the analysis in which the number of iconic gestures in the video context was the criterion variable, there was again a significant effect of experience, $\beta = .45$, $p < .001$, but in addition, scores on the geology tasks predicted iconic gesturing, $p = .001$. Within this third level, both strike and dip accounted for unique variance ($\beta = -.22$, $p = .019$ and $\beta = -.21$, $p = .023$, respectively). Note that the negative beta values reflect the use of a degrees-error measure so that lower scores represent better performance.

4 Discussion

At the most general level, perhaps the most notable finding is the pervasiveness of gestures used by participants in the context of reading about or communicating about the geological concepts of strike and dip. Consistent with the argument that geology—like many other sciences—is a highly spatial domain for which external representations are central, participants made extensive use of iconic gestures. The current data thus add to a growing body of research that has been documenting the role of gestural spatial representations in STEM fields [11, 12, 17, 18, 19, 20, 36, 37]. What is of particular interest, in addition, is the degree to which the use of iconic gestures varied across situational contexts and participant groups. The data provide strong evidence that participants adjust their gesturing dramatically in response to changing task demands, and that the patterns of those adjustments differ in interaction with participant variables.

Gesturing was relatively rare during the instruction-reading phase of the study, and, as expected, it was concentrated in novices. Indeed, the link between experience and this use of gesture was perfect: every one of the participants who gestured during this phase was a novice. We infer that this gesturing is driven by the participant's attempt to understand the material. We found no statistically significant association between gesturing and performance on the subsequent model geology tasks. However, given that gesturing was participant-determined rather than experimenter-manipulated, it is not possible to conclude that gesturing is ineffective for mastering the material. It is possible, for example, that gestures were used by the very participants who were having the most difficulty understanding the material, and thus a failure to find a significant difference in performance between gesturers and non-gesturers might actually indicate a salutatory effect of gesture. Experimental designs are needed in future research to study the impact of gestural training on learning geology just as they have been used to study the impact of gestural training for mental rotation and other educational domains [36, 38, 39].

In contrast to the relatively limited use of gesturing during the instruction-reading phase of the study (16% of novices; 0% of geo-students), there was an overwhelming use of iconic gesture during both audio (70% of novices; 67% of geo-students) and video phases (92% of novices; 100% of geo-students). These data speak to the clear pervasiveness of gesture when communicating about geology to others.

Although at first glance these percentages might appear to suggest that there was only a modest difference across contexts and to imply that there could be little room for variation in relation to participant variables, once the frequencies and loci of gestures are taken into account, it becomes evident that patterns of gesture varied markedly by context and participant characteristics (see Figure 3). Interestingly, among novices and irrespective of WLG, space-locus iconic gestures were more common than either picture-locus or model-locus iconic gestures, both of which were relatively uncommon. There is, though, some indication that these novices did distinguish between the two communicative contexts: in the audio context, gestures were almost exclusively space loci; in the video context, gestures occurred at all loci.

A dramatic finding is the high frequency with which geo-students in the high WLG used gestures involving the model when giving explanations in the video context. Indeed, some geo-students were very explicit about the utility of the model for explaining strike and dip. One, for example, commented favorably about the model, saying "To try to teach a 3-dimensional problem on a 2D surface like a board or something like that, I just think it would be a lot easier if you just had something like this." In contrast, the high WLG geo-students produce virtually no model-directed gestures in the audio context. The fact that these participants gestured with the model when they knew that their intended audience could see their gestures (video context), but did not do so either when their audience could not see them (audio context) or when there was no audience (instruction reading), suggests that for these more spatial and more experienced participants, model-directed gestures are motivated by other-directed (communicative) goals rather than self-directed (cognitive) goals.

Iconic gestures in the surrounding space present a different pattern. Geo-students in both the low and high WLGs produced a relatively high number of iconic gestures in the surrounding space in both audio and video contexts. The presence of space-directed gestures in the audio context, where there is no external audience, suggests that this type of gesture is self-directed rather than communicative. The use of iconic gesturing among these geo-students in the audio context—but not in the instruction-reading phase—may indicate that the communication task is cognitively more demanding, and thus only the former leads them to use gestures as an additional representational support. This finding is reminiscent of the observation that individuals in a challenging geological reasoning task often use 2D or 3D external representations to reduce cognitive load [40].

Another line of evidence bearing on the question of what underlies greater use of gesture comes from the analyses linking strike and dip performance to gestural data. In the instruction-reading phase of the study, novices who gestured did not perform significantly better on strike and dip tasks than those who did not gesture. In the communicative (audio and video) phases of the study, however, the regression analyses did show links between factors associated with better performance on the geology tasks (e.g., geology coursework experience) and greater use of gesture. These findings add support to the notion that those who understand the geological concepts are better positioned to create gestural representations for explaining the concepts to others, much as teachers have been observed to use gesture in science teaching [39, 41]. Future coding of the data collected in the present research will be addressed to examining not only the incidence but also the quality of the gestures (e.g., if hands are held accurately at a horizontal position when reading about strike) and their consistency with spoken language, a central focus of much research on gesture [21, 42, 43]. The incidence and loci of iconic gestures reported here have thus not yet been linked to the number of substantive points explained, their accuracy, their clarity, and so on. In addition to doing more extensive coding and analyses of the data in hand, another important direction for future research is to conduct experimental studies in which the use of gesture is manipulated in both the teacher and student.

Even now, however, the current findings allow us to conclude that some college students find the concepts of strike and dip difficult to understand from reading instructional material, and that gesture is used spontaneously by many students as they seek to understand or communicate these concepts. In addition to expanding our understanding of how students process the spatially-rich discipline of geology, the present research offers a new methodology—manipulating situational context—that may be useful for investigators interested in differentiating between self-directed and other-directed gesture more generally.

Acknowledgments

This research was funded by the National Science Foundation in grants to Liben and Kastens (REC04-11686 and REC04-11823, respectively). All opinions, findings, conclusions, and recommendations are those of the authors and no endorsement from the National Science Foundation should be inferred. This is Lamont-Doherty Earth Observatory Contribution No. 7357. We acknowledge with thanks the contributions of past and current members of the Penn State Cognitive & Social Development Lab, particularly Lauren Pendergast and Daniel Walsh for their help in collecting data. We are grateful to John Sindt of the Lamont-Doherty Earth Observatory Instrument Shop for fabrication of the tabletop models. In addition, we express our appreciation to Susan Goldin-Meadow for her valuable insights about gestural data and coding.

References

1. National Research Council: Learning to think spatially: GIS as a support system in the K-12 curriculum. National Academy Press, Washington (2006)
2. Stieff, M.: Mental rotation and diagrammatic reasoning in science. Learning and Instruction 17, 219–234 (2007)
3. Tversky, B.: Spatial schemas in depictions. In: Gattis, M. (ed.) Spatial schemas and abstract thought, pp. 79–112. MIT Press, Cambridge (2001)
4. Liben, L.S.: Education for spatial thinking. In: Damon, W., Lerner, R. (series eds.), Renninger, K.A., Sigel, I.E. (vol. eds.). Handbook of child psychology, 6th edn. Child psychology in practice, vol. 4, pp. 197–247. Wiley, Hoboken (2006)
5. Kastens, K.A., Rivet, A.: Using analogical mapping to assess the affordances of scale models used in Earth and environmental science education. In: Hölscher, C., et al. (eds.) Spatial Cognition VII. LNCS (LNAI), vol. 6222, pp. 112–124. Springer, Heidelberg (2010)
6. Crowder, E.M., Newman, D.: Telling what they know: The role of gesture and language in children's science explanations. Pragmatics and Cognition 1, 341–376 (1993)
7. Hegarty, M., Mayer, S., Kriz, S., Keehner, M.: The role of gestures in mental animation. Spatial Cognition and Computation 5, 333–356 (2005)
8. Kastens, K.A., Manduca, C.A., Cervato, C., Frodeman, R., Goodwin, C., Liben, L.S., Mogk, D.W., Spangler, T.C., Stillings, N.A., Titus, S.: How geoscientists think and learn. EOS Transactions American Geophysical Union 90, 265–272 (2009)
9. Kastens, K.A., Ishikawa, T.: Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields. In: Manduca, C.A., Mogk, D.W. (eds.) Earth and mind: How geologists think and learn about the Earth, pp. 53–76. The Geological Society of America, Boulder (2006)
10. Kastens, K.A., Rivet, A.: Multiple modes of inquiry in Earth science. The Science Teacher 75, 26–31 (2008)
11. Kastens, K.A., Agrawal, S., Liben, L.S.: Research in science education: The role of gestures in geoscience teaching and learning. Journal of Geoscience Education 54, 262–268 (2008)
12. Singer, M., Radinsky, J., Goldman, S.R.: The role of gesture in meaning construction. Discourse Processes 45, 365–386 (2008)
13. Kastens, K.A., Agrawal, S., Liben, L.S.: How students and field geologists reason in integrating spatial observations from outcrops to visualize a 3-D geological structure. International Journal of Science Education 31, 365–394 (2009)
14. Liben, L.S., Christensen, A.E., Kastens, K.A.: Spatial foundations of science education: The illustrative case of instruction on introductory geological concepts. Cognition and Instruction (forthcoming)
15. Piaget, J., Inhelder, B.: The child's conception of space. Norton, New York (1956)
16. Hostetter, A.B., Alibali, M.W.: Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. Gesture 7, 73–95 (2007)

17. Goodwin, C.: Environmentally coupled gestures. In: Duncan, S., Cassell, J., Levy, E. (eds.) Gesture and dynamic dimensions of language, pp. 195–212. John Benjamins, Amsterdam (2007)
18. Roth, W.-M.: Gestures: The leading edge in literacy development. In: Saul, E. (ed.) Crossing borders in literacy and science instruction: Perspectives on theory and practice, pp. 48–70. International Reading Association, Newark (2004)
19. Roth, W.-M., Lawless, D.: Scientific investigations, metaphorical gestures and the emergence of abstract scientific concepts. Learning and Instruction 12, 285–304 (2002)
20. Broaders, S.C., Cook, S.W., Mitchell, Z., Goldin-Meadow, S.: Making children gesture brings out implicit knowledge and leads to learning. Journal of Experimental Psychology: General 136, 539–550 (2007)
21. Church, R.B., Goldin-Meadow, S.: The mismatch between gesture and speech as an index of transitional knowledge. Cognition 23, 43–71 (1986)
22. Garber, P., Goldin-Meadow, S.: Gesture offers insights into problem-solving in adults and children. Cognitive Science 26, 817–831 (2002)
23. Liben, L.S.: Educational applications of geographic information systems: A developmental psychologist's perspective. In: Barstow, D. (ed.) First national conference on the educational applications of geographic information systems, pp. 44–49. TERC, Cambridge (1995)
24. Liben, L.S., Golbeck, S.L.: Performance on Piagetian horizontality and verticality tasks: Sex related differences in knowledge of relevant physical phenomena. Developmental Psychology 20, 595–606 (1984)
25. Liben, L.S.: Performance on Piagetian spatial tasks as a function of sex, field dependence, and training. Merrill-Palmer Quarterly 24, 97–110 (1978)
26. Thomas, H., Lohaus, A.: Modeling growth and individual differences in spatial tasks. Monographs of the Society for Research in Child Development 58(9), Serial No. 237 (1993)
27. Liben, L.S.: The Piagetian water-level task: Looking beneath the surface. In: Vasta, R. (ed.) Annals of Child Development, vol. 8, pp. 81–143. Jessica Kingsley Publishers, London (1991)
28. Thomas, H., Jamison, W., Hummel, D.D.: Observation is insufficient for discovering that the surface of still water is invariably horizontal. Science 181, 173–174 (1973)
29. Rebelsky, F.: Adult perception of the horizontal. Perceptual and Motor Skills 19, 371–374 (1964)
30. Sholl, M.J., Liben, L.S.: Illusory tilt and Euclidean schemes as factors in performance on the water-level task. Journal of Experimental Psychology: Learning, Memory and Cognition 21, 1624–1638 (1995)
31. De Lisi, R., McGillicuddy-De Lisi, A.V.: Individual differences in adolescents' horizontality representation: Associations with vocational major and gender. Merrill-Palmer Quarterly 34, 437–449 (1988)
32. Shea, D.L., Lubinski, D., Benbow, C.P.: Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. Journal of Educational Psychology 93, 604–614 (2001)
33. U.S. Geologic Survey. (n.d.), `http://geomaps.wr.usgs.gov/parks/gmap/gmap3.html` (Retrieved October 6, 2007)
34. Bakeman, R., Quera, V.: Analyzing interaction: Sequential analysis with SDIS and GSEQ. Cambridge University Press, New York (1995)

35. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics 33, 159–174 (1977)
36. Ehrlich, S.B., Levine, S.C., Goldin-Meadow, S.: The importance of gesture in children's spatial reasoning. Developmental Psychology 42, 1259–1268 (2006)
37. Goldin-Meadow, S., Cook, S.W., Mitchell, Z.A.: Gesturing gives children new ideas about math. Psychological Science 20, 267–272 (2009)
38. Cook, S.W., Mitchell, Z., Goldin-Meadow, S.: Gesturing makes learning last. Cognition 106, 1047–1058 (2008)
39. Valenzeno, L., Alibali, M.W., Klatzky, R.: Teachers' gestures facilitate students' learning: A lesson in symmetry. Contemporary Educational Psychology 28, 187–204 (2003)
40. Kastens, K.A., Liben, L.S., Agrawal, S.: Epistemic actions in science education. In: Freksa, C., Newcombe, N.S., Gärdenfors, P., Wölfl, S. (eds.) Spatial Cognition VI. LNCS (LNAI), vol. 5248, pp. 202–215. Springer, Heidelberg (2008)
41. Alibali, M.W., Flevares, L.M., Goldin-Meadow, S.: Assessing knowledge conveyed in gesture: Do teachers have the upper hand? Journal of Educational Psychology 89, 183–193 (1997)
42. Goldin-Meadow, S., Alibali, M.W., Church, R.B.: Transitions in concept acquisition: Using the hand to read the mind. Psychological Review 100, 279–297 (1993)
43. McNeill, D., Cassell, J., McCullough, K.E.: Communicative effects of speech-mismatched gesture. Research on Language and Social Interaction 27, 223–237 (1994)

Using Analogical Mapping to Assess the Affordances of Scale Models Used in Earth and Environmental Science Education

Kim A. Kastens[1,2] and Ann Rivet[3]

[1] Lamont-Doherty Earth Observatory of Columbia University
[2] Department of Earth & Environmental Sciences, Columbia University
[3] Teachers College, Columbia University
kastens@ldeo.columbia.edu, rivet@tc.columbia.edu

Abstract. Physical analog models are a common pedagogical device in Earth and Environmental Science education for helping students bridge the vast scale difference between the Earth and the classroom. Gentner's structural framework for analogical reasoning has been used to map the correspondences and non-correspondences between two widely-used analog models and the relevant portions of the Earth System. A classroom model of convection in an aquarium has important correspondences to the atmospheric Hadley cell at the levels of attributes, simple relations, higher order relations and systematicity. A volcano eruption model lacks the relations among lava flow temperature, viscosity, and distance that result in construction of the distinctive conical shape of real volcanoes. Analogical mappings of classroom models can be used to guide the design of instruction and assessment so as to increase the chances that students will understand the Earth system at the level of higher-order relations rather than superficial attributes.

Keywords: spatial cognition, analog model, science education, earth science, environmental science.

1 Introduction

Planet Earth is 18 orders of magnitude larger by volume than a typical classroom. The Earth's environment, taken as the envelope of life-supporting air and water that surrounds our rocky planet, is approximately sixteen orders of magnitude larger than a classroom. Even a local environment, for example a smallish watershed, may be four or five orders of magnitude larger than a classroom by area.

This inalterable fact of scale profoundly influences the teaching of Earth and Environmental Sciences, and sets these disciplines apart from other school sciences. In Life Sciences, students can grow plants on the windowsill, actively photosynthesizing and respiring. In Physical Sciences, the teacher can bring an acid or lever or electric circuit into the classroom, and students can create an authentic chemical reaction or force or electrical current.

But neither students nor teacher can create an authentic earthquake or hurricane or ocean current in the classroom. In fact, most of the important phenomena and structures

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 112–124, 2010.

that feature in Earth Science or Environmental Science curricula cannot fit within the four walls of a classroom. As a consequence, many of the laboratory activities in these courses are analogies for the phenomena that the curriculum is targeting, rather than the real thing. Rather than manipulate the actual phenomenon under study, gathering evidence of its existence and behavior, students manipulate an analog model.

Curriculum developers have created an impressive array of such models, which are fun, affordable, safe, and suited for use by primary and secondary students (Fig. 1). However, there is little educational research on what or how students learn *about the Earth* from such models. Although students find hands-on activities with analog models engaging, it seems possible that some students are learning about the model but not successfully transferring insights from the classroom-scale model to the full scale Earth System.

As one facet of a comprehensive study of the use of analog models in Earth Science instruction, we are developing methodology to evaluate the affordances of individual classroom models. "Affordances," in this context, refers to learnings that the model is capable of supporting. No analogy makes a perfect mapping between source domain and target domain, and this is true of analog models used in teaching. Any given model has the potential to convey some aspects of how the Earth works, and will fail to convey other aspects.

The nature of the model itself is only one factor that influences the effectiveness of model-supported teaching and learning. The pedagogical context provided by the teacher and curriculum materials is critically important, such as guiding questions, diagrams, and opportunities for inquiry and discourse. Some attributes of the students are surely important as well, such as their prior knowledge of the Earth phenomena under study and their prior experience learning from analog models. But an upper boundary for educational effectiveness of any specific model is set by the nature of the model itself, the focus of this paper.

2 Analogical Mapping from Classroom Model to Earth System

Gentner's structure mapping framework for analogy ([7], [8], [9], [10]) offers a robust theoretical framework within which to analyze each specific analog model and against which to calibrate both learning goals and student performance. Gentner's framework distinguishes among different forms of similarity that may exist between the source and the target, and articulates a set of implicit rules for mapping knowledge about the source onto the target. This body of work shows that the power of analogy comes from the relationships between objects, rather than from the attributes of objects, and that the most powerful analogy-derived insights come from the existence of higher-order relations such as causality.

Using Gentner's framework requires parsing the behavior of the model system and the relevant portion of the Earth System into statements in the form of "predicates," which are verb phrases (shown in all caps) with one or more "arguments" (enclosed in brackets or parentheses). "Arguments" can be either objects or propositions, and they can be either the subject or the object of the verb phrase. The predicate can contain an action verb, like RISES(water), or a verb of state, like IS-TRANSPARENT(water). Parsing the behavior of the two systems in this manner makes explicit the correspondences and non-correspondences between the model and Earth. Gentner's framework specifies four levels of similarity:

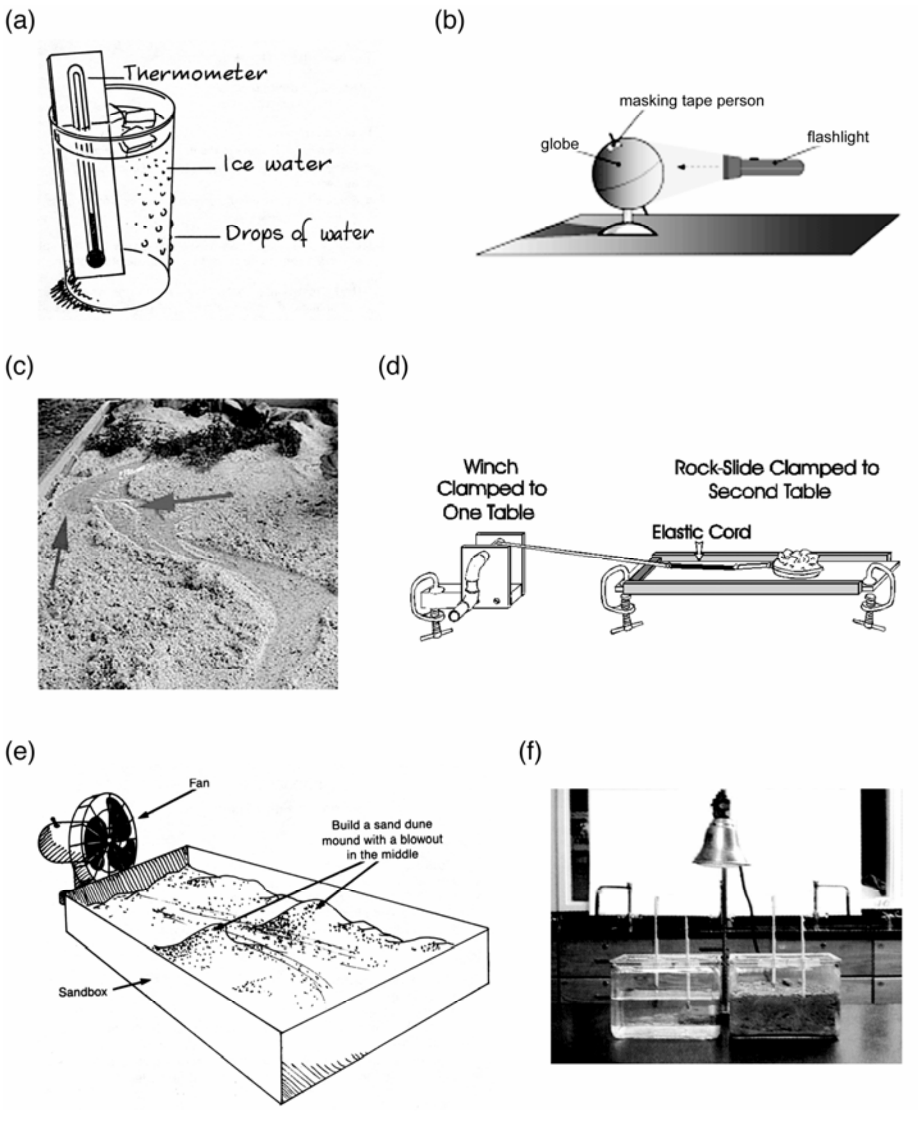


Fig. 1. Examples of analog models used in teaching Earth and Environmental Science. (a) Dew point model [1]. (b) Earth's rotation model [2]. (c) Stream table model of erosion and deposition [3]. (d) Model of stick-slip behavior of earthquake faults [4]. (e) Sand dune model [5]. (f) Model of differential heating of ocean and continents [6].

- *Attributes:* predicates have only one argument and the argument is an object.
- *Simple relations:* predicates have two or more arguments and the arguments are objects.
- *Higher order relations:* the arguments are propositions rather than objects.
- *Systematicity:* predicates include deeply nested relational chains.

We now analyze two widely used Earth Science teaching models and the relevant portion of the Earth System, according to this framework. The examples chosen rely heavily on spatial thinking, in that aspects of the Earth System illustrated by these classroom models concern shape, configuration, and trajectory.

2.1 Model #1: Convection in an Aquarium

Figure 2(a) shows a model described by Freeman [11], which he has used in high school Earth Science and Oceanography courses. A water-filled aquarium is elevated above the table surface such that a row of small candles can be inserted beneath the aquarium forming a line across the short dimension of the aquarium. Copper pennies are used to create a trough along the floor of the aquarium above the candles, and a stripe of red food coloring is carefully laid down in the trough. Blue food coloring is placed near the two ends of the aquarium at the top of the water column. The candles are lit, and gradually the heated water rises, forming the upwelling limb of a convection cell, marked with tendrils of red color. Slightly later, the downwelling limbs of the convection cell become apparent, marked with tendrils of blue. Variants of this model are common in laboratory manuals and teachers' guides; variations include cooling the upper water with ice rather than heating the lower water, and achieving the density contrast with salinity rather than temperature.

Figures 2(b) and 2(c) are textbook illustrations of portions of the Earth System that are considered to be analogous to the model of figure 2(a). Figure 2(b) shows air rising over the equator, traveling northward and southward away from the equator at high altitude, sinking at 30°N and S latitude, and returning along the earth's surface towards the equator, in a large scale atmospheric circulation cell called the Hadley cell. Figure 2(c) shows mantle material rising beneath a mid-ocean ridge spreading center, lithosphere translating laterally in both directions away from the spreading center, downwelling beneath subduction zones, and a return flow at depth back towards the spreading center. Both textbook diagrams are greatly simplified from the actual Earth System, but the diagrams represent the level of understanding expected of high school or introductory level college students.

Figure 3 shows the analogical mapping between the model system and the Hadley cell portion of the Earth's atmospheric circulation system. Important correspondences at the *attribute* level are RISES(water/air), SINKS(water/air), IS-FLUID(water/air), and IS-WARM(candle/solid-earth). An important non-correspondence at the attribute level concerns scale: the convection cells in the classroom model are of centimeter scale, while the Hadley cells are of kilometer scale.

The attribute-level description of the model system includes many of the most obvious aspects of the system. If left to their own devices to observe and describe the model system, students are most likely to observe attribute-level aspects of the model system, for example the rising water, the sinking water, and the hot candle. Unfortunately, attribute-level aspects of the model system may not have much explanatory power. Some conspicuous attributes of the model system have no correspondent in the Earth system. For example, IS-RED(water) and IS-BLUE (water) have no counterpart in the atmosphere. Even in cases where the source system and the target system have similar or identical predicates, this may carry little explanatory or predictive power, for example IS-TRANSPARENT.

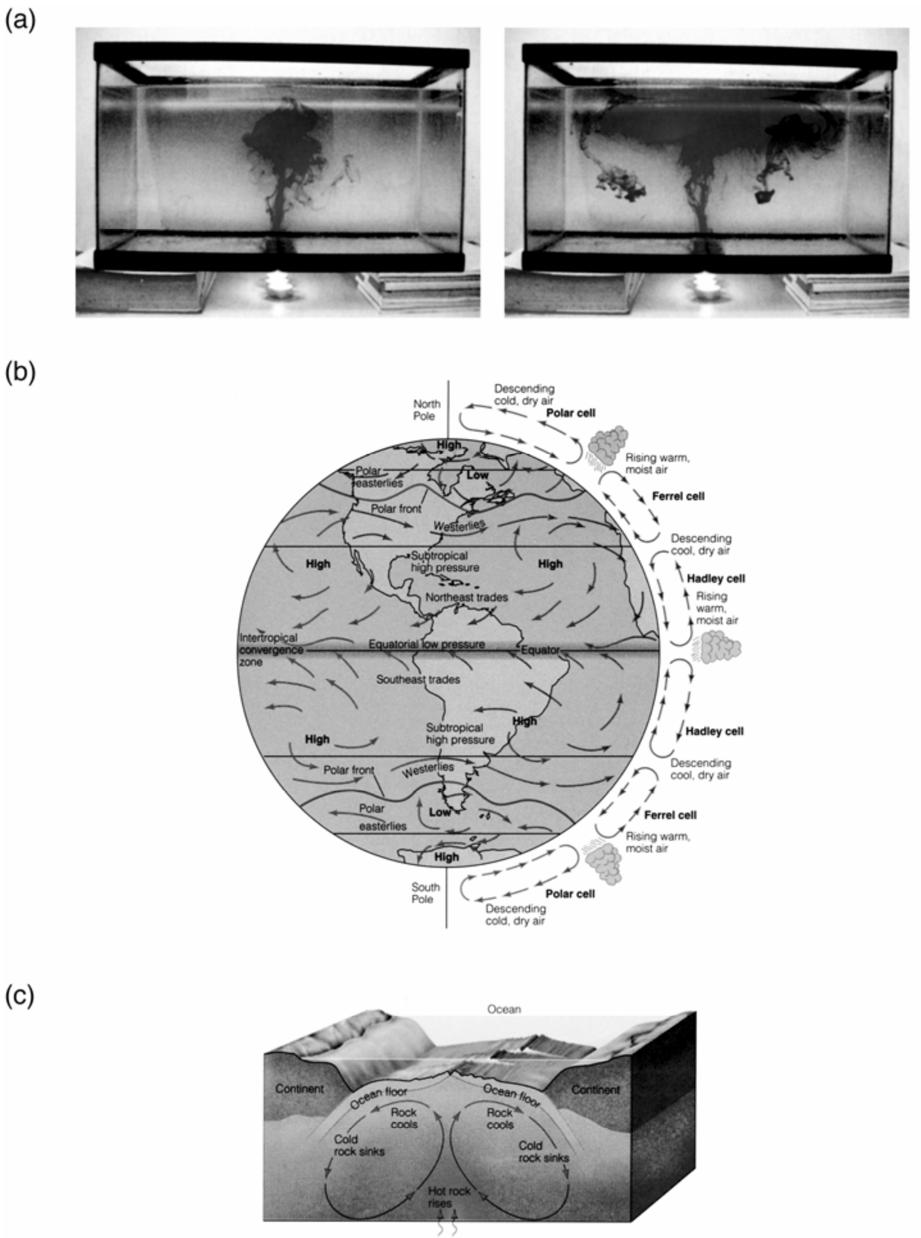


Fig. 2. (a) "Convection in a Fishtank" model [11], recommended for use in Earth Science and Oceanography courses at the high school level. The upwelling limb of the convection cell is red and the downwelling limbs are blue. (b) Textbook illustration of convection cells in the atmosphere [12, p. 301]. The best developed convection cells are the pair centered on the equator, known as "Hadley cells." (c) Textbook illustration of convection cells in the Earth's mantle [12, p. 68].

Aquarium convection: tabletop system **(source domain of analogy)**		***Hadley Cell: full-scale Earth System*** **(target domain of analogy)**
Attributes: predicates have only one argument and the argument is an object		
• RISES(water)	⇔	• RISES(air)
• SINKS(water)	⇔	• SINKS(air)
• IS-WARM(candle)	⇔	• IS-WARM(solid-earth)
• IS-RED(water)	⊗	
• IS-BLUE(water)	⊗	
• IS-TRANSPARENT(water)	⇔	• IS-TRANSPARENT(air)
• IS-LIQUID(water)	⊗	• IS-GASEOUS(air)
• IS-FLUID(water)	⇔	• IS-FLUID(air)
• IS-CM-SCALE(convection cell)	⊗	• IS-KM-SCALE(convection cell)
Simple relations: predicates have two or more arguments and the arguments are objects		
• IS-UNDERNEATH(lower-water, upper-water)	⇔	• IS-UNDERNEATH(lower-air, upper-air)
• RISES-AND-DISPLACES(lower-water, upper-water)	⇔	• RISES-AND-DISPLACES(lower-air, upper-air)
• PROVIDES –HEAT-TO(candle, lower-water)	⇔	• PROVIDES-HEAT-TO(solid-earth, lower-air)
• IS-LESS-DENSE-THAN(lower-water, upper-water)	⇔	• IS-LESS-DENSE-THAN(lower-air, upper-air)
Higher order relations: the arguments are propositions rather than objects		
• CAUSES[PROVIDES-HEAT-TO(candle, lower-water), IS-LESS-DENSE-THAN(lower-water, upper-water)]	⇔	• CAUSES[PROVIDES-HEAT-TO(solid-earth, lower-air), IS- LESS-DENSE -THAN(lower-air, upper-air)]
• CAUSES[IS- LESS-DENSE -THAN(lower-water, upper-water), RISES&DISPLACES(lower-water, upper-water)]	⇔	• CAUSES[IS- LESS-DENSE -THAN(lower-air, upper-air), RISES&DISPLACES(lower-air, upper-air)]
Systematicity: deeply nested relational chains		
• CAUSES{PROVIDES-HEAT-TO(candle, lower-water), CAUSES[IS-LESS-DENSE -THAN(lower-water, upper-water), RISES&DISPLACES(lower-water, upper-water)]}	⇔	• CAUSES{PROVIDES-HEAT-TO(solid-earth, lower-air), CAUSES[IS-LESS-DENSE –THAN (lower-air, upper-air), RISES&DISPLACES(lower-air, upper-air)]}

Fig. 3. Analogical mapping between the aquarium convection model and the atmospheric Hadley cell. In the center column, double-headed arrows mark correspondences between the classroom model system and the Earth system; circled X's mark aspects of the source and target that do not correspond.

Simple relations record a relationship between the first argument and the second argument. The relationship can be a comparison, with no action implied, i.e. the lower-water IS-LESS-DENSE-THAN the upper-water. Or the relationship can be that the first argument acts on the second argument, i.e. the candle PROVIDES-HEAT-TO the lower-water, or the lower-water RISES-AND-DISPLACES the upper-water. The power of the aquarium model begins to emerge at this level of analogy, where we see that the aquarium and the Hadley cell have exactly parallel constructions in Figure 3, except that the fluid is "water" in the aquarium and "air" in the atmosphere. Understanding the system at this level requires mentally breaking the "water" object into parcels of water ("upper-water" and "lower-water") that can be acted upon separately and can behave differently from each other, even though there is no concrete boundary between them. Whereas the attribute-level aspects of the aquarium model can be seen by the observant eye, understanding the system at the level of simple relations requires application of learned concepts: density and heat.

At the level of *high order relations,* predicates begin to capture causality. The candle PROVIDES-HEAT-TO the lower-water, which CAUSES lower-water to BE-LESS-DENSE-THAN than upper-water. The lower-water IS-LESS-DENSE-THAN the upper-water, which CAUSES lower-water to RISE&DISPLACE upper-water. At this level, also, we see parallel structures between the aquarium model system and the Hadley cell system. Note that the higher order relations of Figure 3 are composited from simple relations; thus this level cannot be understood until the simpler levels are understood.

The *systematicity* level nests all of the relations from the previous two levels. Unraveling the propositions in the aquarium convection system gives us: candle PROVIDES-HEAT-TO lower water, which CAUSES lower-water TO-BE-LESS-DENSE THAN upper-water, which CAUSES lower-water to RISE-AND-DISPLACE upper-water. Exactly the same structure is found in the Hadley cell convection system. Thinking about the system at this level can generate profound questions about the workings of the Earth System. For example, it is obvious by inspection that the candles PROVIDE-HEAT-TO the lower-water in the aquarium. If the rising column of air over the equator is to be considered analogous to the rising column of water above the candles, then it must be the case that the atmosphere is heated from below. Yet students know that the sun is the source of heat in the atmosphere, and the sun shines down onto the top of the atmosphere, not up from the bottom like the candle in the classroom model. Confronting this discrepancy can lead students into a deeper understanding of energy, light, heat and earth processes, as they put together the understanding that the atmosphere is actually heated from below by heat re-radiated upward from the solid earth and oceans.

In summary, the aquarium convection model and the Hadley cell have important correspondences at the attribute, simple relations, higher order relations and systematicity levels. Thus the aquarium convection model affords the possibility of supporting deep learning about convection cells in natural systems. On the other hand, conspicuous non-correspondences at the attribute level may present obstacles to learning.

2.2 Model #2: Baking Soda Volcano

Figure 4 illustrates one of the most common Earth Science activities in collections of "hands-on" science activities for elementary and middle school children. A conical structure is built up of paper maché or other crafts material around a can or bottle, leaving the opening of the can or bottle exposed at the apex of the cone. Vinegar, or a mixture of vinegar and water is poured into the bottle, and a few drops of red and yellow food coloring are added. Then a few tablespoons of baking soda are dropped into the vinegar mixture. The vinegar and baking soda react violently to form carbon dioxide, and the mixture of gas and liquid comes bubbling out the apex of the cone and flows down the sides in a manner suggestive of lava flowing from a volcano. Children tend to love this activity, whereas professional geoscientists tend to scorn it as superficial and irrelevant to real world volcanic processes.

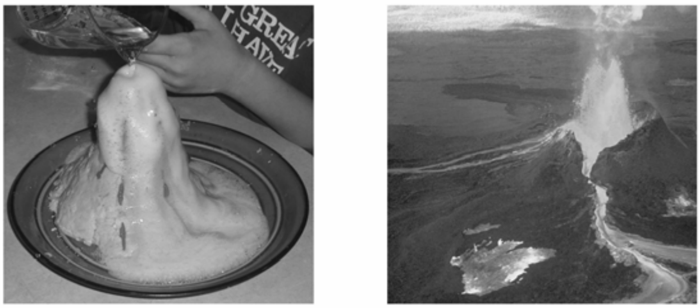


Fig. 4. (a) "Baking soda volcano" model. This illustration is from Helmenstine [13], but variations of this activity are common in activity books and teachers' guides for elementary and middle school aged children. (b) Photograph of an authentic volcanic eruption [14].

Figure 5 shows our analogical mapping between the baking soda volcano and a real world erupting volcano. The Earth System end of the mapping focuses only on the behavior of lava flows, leaving aside ash falls, volcanic mudflows (lahars), and other processes that may be prominent in specific volcanoes. As with the aquarium convection model, an important *attribute* level non-correspondence between classroom model and Earth system concerns the matter of scale: the classroom model IS-CM-SCALE whereas the authentic volcano IS-KM-SCALE. As with the aquarium model, many of the most conspicuous aspects of the baking soda volcano model are attributes, notably the conical shape of the structure, and the behavior of the effluent as it emerges suddenly and voluminously from the apex of the structure and flows down the sides. These model attributes correspond reasonably well to the attributes of authentic volcanoes and eruptions.

One conspicuous non-correspondence at the *attribute* level--that lava IS-HOT while the vinegar mixture IS-NOT-HOT--might appear at first glance to be merely a matter of safety and convenience, but deeper in the analysis this emerges as an important causal factor. At the simple relations level, a major non-correspondence is that the volcano model was build by humans out of crafts materials, whereas the authentic volcano was built up out of solidified effluent, i.e. lava. Stepping down through the levels of the analysis, we can see how the distinctive conical shape of the authentic volcano builds up, beginning from the attribute that lava IS-HOT. Because hot magma emerges into cooler air, the authentic volcano HAS-TEMPERATURE-GRADIENT from hotter,

Baking Soda Volcano: tabletop system **(source domain of analogy)**		***Volcanic eruption: full-scale Earth System*** **(target domain of analogy)**
Attributes: predicates have only one argument and the argument is an object		
• IS-CONICAL(structure)	⇔	• IS-CONICAL(structure)
• EMERGES-FROM-APEX(effluent)	⇔	• EMERGES-FROM-APEX(effluent)
• EMERGES-SUDDENLY-AND-VOLUMINOUSLY(effluent)	⇔	• EMERGES-SUDDENLY-AND-VOLUMINOUSLY(effluent)
• FLOWS-DOWN-CONE-SIDES(effluent)	⇔	• FLOWS-DOWN-CONE-SIDES(effluent)
• IS-NOT-HOT(effluent)	⊗	• IS-HOT(effluent)
• IS-CM-SCALE(structure)	⊗	• IS-KM-SCALE(structure)
Simple relations: predicates have two or more arguments and the arguments are objects		
• IS-MIXTURE-OF(effluent, liquid water & carbon dioxide gas)	⇔	• IS-MIXTURE-OF(effluent, liquid magma & gases & solids)
• ARE-REACTION-PRODUCT-OF(gas bubbles, vinegar + baking soda)	⊗	• EXSOLVE-FROM(gas bubbles, magma)
• CAUSE-OVERPRESSURE(gas bubbles, chamber)	⇔	• CAUSE-OVERPRESSURE(gas bubbles, chamber)
• WAS-BUILT-BY(structure, humans)	⊗	• WAS-BUILT-BY(structure, solidified effluent)
	⊗	• HAS-TEMPERATURE-GRADIENT(recent effluent, older effluent)
	⊗	• IS-MORE-VISCOUS-THAN(cooler effluent, hotter effluent)
Higher order relations: the arguments are propositions rather than objects		
• CAUSES[IS-OVERPRESSURED(gas bubbles in chamber), EMERGES-SUDDENLY-AND-VOLUMINOUSLY(effluent)]	⇔	• CAUSES[IS-OVERPRESSURED(gas bubbles in chamber, EMERGES-SUDDENLY-AND-VOLUMINOUSLY(effluent)]
• HAS-POTENTIAL-ENERGY-GRADIENT[TOP(IS-CONICAL(structure), BOTTOM(IS-CONICAL(structure)]	⇔	• HAS-POTENTIAL-ENERGY-GRADIENT[TOP(IS-CONICAL(structure), BOTTOM(IS-CONICAL(structure)]
	⊗	• CAUSES-VISCOSITY-GRADIENT(HAS-TEMPERATURE-GRADIENT(recent effluent, older effluent), IS-MORE-VISCOUS-THAN(cooler effluent, hotter effluent)]
Systematicity: deeply nested relational chains		
• CAUSES{POTENTIAL-ENERGY-GRADIENT[…], FLOWS-DOWN-CONE-SIDES(effluent)}	⇔	• CAUSES{POTENTIAL-ENERGY-GRADIENT[…], FLOWS-DOWN-CONE-SIDES(effluent)}
• DETERMINES-STOP-POINT{POTENTIAL-ENERGY-GRADIENT[BOTTOM(IS-CONICAL(structure)], FLOWS-DOWN-CONE-SIDES(effluent)}	⊗	• DETERMINES-STOP-POINT{VISCOSITY-GRADIENT[….], FLOWS-DOWN-CONE-SIDES(effluent)}

Fig. 5. Analogical mapping between the baking soda volcano model and an erupting volcano in the Earth system

more recent effluent near the vent opening to cooler, older effluent farther away from the vent opening. This important *simple relation* in the authentic volcano has no counterpart in the model volcano. Lava becomes more viscous as it cools, and thus at the *higher order relations* level of the authentic volcano we see that HAS-TEMPERTURE-GRADIENT CAUSES HAS-VISCOSITY-GRADIENT between hotter, less viscous, newly-erupted lava near the vent opening and cooler, more viscous lava farther from the vent. Again, there is no counterpart relation in the classroom model. At the *systematicity* level, we see that in both the classroom model and the authentic volcano, the behavior of FLOWS-DOWN-CONE-SIDES is initiated and maintained because the conical structure HAS-POTENTIAL-ENERGY-GRADIENT between the top and bottom of the structure. Although the causality of initiating and maintaining the flows in the model system and Earth system are the same, the causality of stopping the flows is different. In the classroom model, with its low viscosity vinegar mixture, the STOP-POINT of the flow occurs when the effluent runs out of potential energy at the base of the structure and puddles in the tray or spills onto the table. With a real lava flow on a real volcano, STOP-POINT of each flow occurs when the cooling lava becomes too viscous for continued motion on the available slope, and freezes on the spot. The distinctive IS-CONICAL shape of an authentic volcano is an emergent property caused by successive lava flows, each traveling part of the way down the volcano's flank and freezing in place, with larger numbers of flows traveling shorter distances.

Although the baking soda volcano model does not afford an understanding of the causal relationships that underlie the distinctive conical volcano shape, there is another aspect of eruption behavior that could be supported with this model. At the *simple relations* level, we note that the effluent in both the model and the volcano IS-MIXTURE-OF liquid and gas. Although the source of the gas does not correspond, in both systems the presence of gas bubbles in the mixture CAUSES-OVERPRESSURE in the chamber. At the *higher relations* level, this overpressure CAUSES the effluent to EMERGE-SUDDENLY-AND-VOLUMINOUSLY. Thus a potential affordance of this model would be to support learning about the role of gas in building up the pressure that causes eruptions. However, the model is rarely used to teach this concept. At the grade levels where this model is typically used, magma is usually defined as "molten rock"; gas is not even mentioned. It seems possible that the familiar baking soda volcano could be reintroduced after students have some understanding of the relations among pressure, temperature, and changes of state from their studies of physical sciences. At this point, the baking soda volcano might be used to trigger an inquiry about whether there is gas in magma (answer: yes), where the gas comes from (answer: it exsolves from the liquid), and what makes the gas form (answer: as the magma rises, the pressure decreases, enabling the change of state of the most volatile components of the magma from liquid to gas).

In summary, the baking soda volcano model lacks the crucial set of relations between the temperature and viscosity of the effluent that cause authentic lava to cease flowing and freeze in place in the course of moving away from the vent. The low viscosity vinegar mixture simply flows down the potential energy gradient of the sloping side of the cone until it reaches the flat tray or table, and does not contribute to building the structure. This non-correspondence undermines one of the most important meta-messages of Earth Science education: that the form of extant Earth structures preserves clues to their formative processes. On the other hand, the model does afford an understanding of the role of gas in causing the overpressure that leads magmatic eruptions to be sudden and effusive.

3 Discussion

Analogies are common in geoscience education, perhaps because so much of the field deals with processes and forces that students cannot perceive directly [15]. In addition to the physical models emphasized here, earth science educators and curriculum materials make abundant use of verbal and diagrammatic analogies [16].

Gentner's structural framework for analogical reasoning has proven to be a powerful methodology for revealing and documenting both the linkages and disconnects between two widely used classroom teaching models and the analogous phenomena in the full-scale Earth System. Application of this approach to the aquarium convection model confirms teachers' reports that this model has the potential to support deep insights about the analogous Earth processes, as shown by the presence of mappings at the level of higher order relations and systematicity. Application of this approach to the baking soda volcano model confirms the criticism that the model fails to support understanding of the relationship between the shape of the volcano and its formative process. However, the mapping also revealed an underappreciated and underexploited source-target correspondence concerning the role of pressurized gas in causing the sudden and voluminous effusion from both model and volcano.

The two systems analyzed share an important non-correspondence at the attribute level, in that the classroom model is at the scale of centimeters while the analogous portion of the Earth System is at the scale of kilometers. This non-correspondence is common to almost all Earth Science teaching models, and is an inevitable consequence of the discrepancy between the size of the Earth and classroom. One exception would be the laboratory activity in which crystals of a low-melting temperature material are grown at room temperature (forming large crystals) and on ice (forming small crystals), which is a classroom analog for slowly cooling intrusive and quickly-cooling extrusive igneous rocks [17]. In that model, the length scale is similar between classroom model and Earth system, but the temporal scale is different.

In both of the systems analyzed, the features of the classroom model most likely to be noticed by students left on their own to explore, observe and describe the model are features at the attribute level. These include the rising and sinking water of the aquarium convection model, and the sudden and voluminous emergence of effluent in the baking soda volcano model. Purposefully crafted curriculum materials and skillful guidance from the teacher [18] will be needed to ensure that students notice and understand what is going on in the classroom model at the level of relations or systematicity. If students have not understood these higher order relations in the model system, the analogy cannot support students in understanding the higher order relations in the Earth system.

When utilizing such models as part of science instruction, some of the most profound learning opportunities arise if and when students critically examine the correspondences and non-correspondences between the classroom model and the Earth system. If students question "what would be heating the atmosphere from below like the candles in the model?" or "what in Nature builds up the pressure like the gas bubbles in the model?", the door is opened to a deeper understanding of Earth processes. Note that the model does not convey this deeper understanding by itself. Excellent inquiry-based pedagogy is required to bring students to the point where they will notice the potential discrepancy and pose the critical questions, and then to guide them to find answers to their questions through reasoning and research. Kurtz, et al [19] have shown that merely exposing students to an analogous pair of situations side by side is of limited effectiveness in fostering their ability to detect meaningful differences. Substantial insights about similarities and differences between the analogs emerged only when the experimenters required their undergraduate participants to articulate the analogical mapping by assigning specific correspondences between elements of the two scenarios in a written exercise. The bottom line is that *both* high-quality models and high-quality instruction are needed to most effectively support students in developing their understanding of the Earth system.

We foresee three applications of this line of work. First, would be to give guidance to teachers and curriculum developers regarding their choices of classroom physical models, favoring models that have more substantial affordances. Second, would be to inform the design of learning activities and assessments, to ensure that these materials foster and probe students' understanding of higher order relations and systematicity in the model and the Earth System, not just the more superficial attributes and simple relations. Third, would be to develop teachers' guides and professional development training to help teachers perceive and convey the both the simpler and more complex correspondences between model and Earth systems. We do not anticipate that most teachers will be able to work directly from analyses such as figures 3 and 5; we will be testing other representational strategies such as concept maps to convey the gist of the analogical mapping in a more user-friendly format. In the remaining years of our project, we will be carrying out methodical observation and analysis of teachers' and students' use of tabletop models in middle- and high-school Earth Science classrooms, with the goal of understanding and improving how students learn about the Earth System through exploration of analog models.

Acknowledgements. The authors thank Dedre Gentner and her research group for illuminating discussions, participants in Spatial Cognition 2008 for introducing KK to the structural framework for analogical reasoning, L. Pistolesi for preparing the manuscript and illustrations, and the National Science Foundation for support through grant DRL09-09982. The opinions are those of the authors and no endorsement by NSF is implied. This is Lamont-Doherty Earth Observatory contribution number 7356.

References

1. Van Cleave, J.: Earth Science for Every Kid. John Wiley & Sons, New York (1991)
2. NASA's Kids Science News Network^{TM}: What causes night and day?, http://ksnn.larc.nasa.gov/night_and_day.html
3. Bryant Watershed Education Project: A Stream Table Makes a Miniature Stream, http://www.watersheds.org/earth/streamtable.htm
4. Lahr, J.C.: Table-top Earthquakes, http://jclahr.com/science/earth_science/tabletop/index.html
5. Bonnet, R.L., Keen, G.D.: Earth Science: 49 Science Fair Projects. McGraw-Hill, Inc. (TAB Books Division), New York (1990)
6. Feather Jr., R.M., Snyder, S.L., Hesser, D.T.: Merrill Earth Science. Macmillan/McGraw-Hill (Glencoe Division), Westerville (1993)

7. Gentner, D.: Structure-mapping: A theoretical framework for analogy. Cog. Sci. 7, 155–170 (1983)
8. Gentner, D., Gentner, D.R.: Flowing waters or teeming crowds: mental models of electricity. In: Gentner, D., Stevens, A.L. (eds.) Mental Models, pp. 99–129. Lawrence Erlbaum, Hillsdale (1983)
9. Gentner, D., Toupin, C.: Systemicity and surface similarity in the development of analogy. Cog. Sci. 10, 277–300 (1986)
10. Gentner, D., Colhoun, J.: Analogical processes in human thinking and learning. In: von Müller, A., Pöppel, E. (Series eds.), Glatzeder, B., Goel, V., von Müller, A. (vol. eds.) On Thinking: Towards a Theory of Thinking, vol. 2, pp. 1148–1158. Springer, Heidelberg (2010)
11. Freeman, C.: Convection in a Fish Tank. The Science Teacher 75, 62–66 (2008)
12. Skinner, B.J., Porter, S.C., Botkin, D.B.: The Blue Planet. John Wiley & Sons, Inc., New York (1999)
13. Helmenstine, A.: How to build a baking soda volcano, `http://chemistry.about.com/cs/howtos/ht/buildavolcano.htm`
14. USGS Volcano Hazards Program, `http://volcanoes.usgs.gov/images/pglossary/effusive.php`
15. Jee, B.D., Uttal, D.H., Gentner, D., Manduca, C., Shipley, T.F., Tikoff, B., Ormand, C.J., Sageman, B.: Commentary: Analogical Thinking in Geoscience Education. Journal of Geoscience Education 58(1), 2–13 (2010)
16. Sibley, D.: A cognitive framework for reasoning with scientific models. Journal of Geoscience Education 57(4), 255–263 (2009)
17. Illinois State Museum Geology Online, `http://geologyonline.museum.state.il.us/tools/lessons/12.1/index.html`
18. Richland, L.E., Zur, O., Holyoak, K.J.: Cognitive supports for analogies in the mathematics classroom. Science 316, 1128–1129 (2007)
19. Kurtz, K.J., Miao, C.-H., Gentner, D.: Learning by analogical bootstrapping. Journal of the Learning Sciences 10(4), 417–446 (2001)

Aligning Spatial Perspective in Route Descriptions

Elena Andonova

SFB/TR8 Spatial Cognition, 5, Enrique Schmidt Str., Cartesium,
University of Bremen, Bremen, Germany
andonova@uni-bremen.de

Abstract. Spatial perspective refers to the use of reference systems in extended spatial descriptions and wayfinding. Variable viewpoints, conceptualizations, and reference terms may lead speakers to describe an environment and movement in it in a route perspective or a survey/gaze perspective. Previous research has indicated the role of a number of environmental, individual, and learning factors in choice of perspective. Here, in two experiments set up in a confederate experimental paradigm with speakers taking turns in describing routes on schematic maps, choice of spatial perspective was influenced also by the use of perspective of the dialogic partner, both before and after they switch perspective. Perspective priming did not occur when partners used perspective inconsistently, however. Participants with low spatial ability as assessed in a self-report measure did not align with their partner when the switch was from route to survey perspective. Results are discussed in the framework of interactive alignment. These studies also found considerable inter-individual variation in the stability vs. flexibility of spatial perspective use.

1 Introduction

When attempting to describe how to get from a departure point to a destination, one can employ a multitude of means, from simple pointing to entire novels. One and the same entity or event in reality may be associated with various alternative representations in our mind, at the level of percepts, concepts, and verbal expression. Such variety exists not only within an individual's cognitive representation but can also be revealing of differences across minds. Communication about the world is grounded in schemas which are mutually shared, thus mutually comprehensible, and at the same time also flexible and adaptive to the situation at hand and to the person we try to communicate with.

In spatial reference, different conceptualizations can be found in the choices of spatial perspective and frames of reference. In particular, perspective taking involves abstracting from the visual scene and organizing and packaging information in accordance with one or another type of viewpoint. Spatial perspective varieties can be characterized in different ways. Here we will adopt a binary distinction which is a simplified yet common typology [1, 2]. A route or environment can be described from an embedded (route or egocentric) perspective, that is, from within the environment, based on the way-finder, as embedded in the path, or from an external (survey or allocentric) perspective, that is, a viewpoint external to the environment, commonly

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 125–138, 2010.

associated with maps and cardinal directions, the way people would look at a map or a drawing of a route. For the sake of brevity and simplicity, we will refer to these as route perspective and survey perspective, a distinction that is examined in the research reported here. Previous studies have demonstrated that a number of individual, environmental, and learning factors can produce variation in spatial perspective in verbal descriptions. Mode of acquisition has been shown to affect perspective choices in spatial memory, for example, participants who studied maps gave more accurate responses later in survey perspective tasks whereas participants who were doing navigation gave more accurate responses to route perspective tasks [3]. Taylor & Tversky [4] tested the influence of four environmental features on spatial perspective choices and found that although overall most participants' descriptions followed a survey or a mixed perspective, preference for the use of route perspective was enhanced in environments that contained a single path vs. multiple paths and environments that contained landmarks of a single size scale vs. landmarks of varying size. Bugman, Coventry, and Newstead [5] found that context of retrieval (frequency of visitation vs. importance of activities) can affect spatial perspective choices, too.

Intra-individual variability in spatial perspective choices involves speakers switching perspective – participants tend to mix perspectives quite regularly, for example, 27 out of 67 participants in Taylor & Tversky's [4] first experiment and 74 out of 192 participants in their second experiment mixed perspectives in their descriptions. On the other hand, consistency in the use of a reference frame has also been established in a number of studies, for example, Vorwerg [6] found that speakers tended to repeat spatial reference frame, lexical and syntactic choices across successive spatial utterances in a localization task. There are multiple reasons why a speaker may or may not switch from one perspective to another, for instance, because of some features of the environment or the task. However, although most studies have researched spatial perspective choices in a monologue setting, one important reason for initial perspective choice and subsequent switches may be the behavior of the interlocutor (conversation partner) in a typical dialogue setting of giving road instructions, for example. In recent years, in addition to the dominant monologue settings, several studies have attempted to address the issue of communicative interaction. Schober [7] showed that speakers set spatial perspectives differently with actual addressees than with imaginary ones. In a study by Striegnitz, Tepper, Lovett, & Cassel [8] there was an increased use of survey perspective in response to clarification questions and in rephrasal of previously given route descriptions. Watson, Pickering, & Branigan [9] found priming effects of reference frame in a confederate task where speakers described the location of an abstract shape with respect different objects on a series of drawings.

Of particular interest and relevance to the research reported here are two studies that reveal the importance of individual differences. Levelt [10] found evidence for differences in what he called 'cognitive styles' between 'jumpers' and 'movers' in spatial descriptions of a network of abstract nodes, and between cognitive types such as 'pattern-oriented' and 'ego-oriented' types. Participants in this study showed a rather consistent style of orientation in describing spatial patterns, both within and across descriptions of a series of 53 patterns. In a task where speakers (directors) described spatial arrangements to their partners (matchers), Schober [11] tested for the influence of individual spatial ability on choice of perspective (with respect to viewpoint) and how successful the interaction of the two partners in a dyad was in terms of comprehension accuracy. In his study, low ability directors were more likely to take their own perspective while high ability directors were more likely to take their partner's perspective. Ability did not affect the use of neutral and object-centered descriptions. Generally, the literature on spatial cognition, and in particular, on wayfinding and route descriptions, abounds in examples of variability in skills and preferences across individuals. To what extent individual spatial ability may affect speakers' descriptive behavior in terms of consistency and flexibility of use of spatial perspective is still an open question.

The variability of spatial perspective and perspective switching also make this phenomenon a suitable testing ground of coordination of speakers' choices in dialogue. Thus, two strands of research and related questions are in the combined focus of this paper—spatial perspective use and dialogic interaction.

The interactive alignment model [12] posits that a large share of speech production choices in dialogic situations can be explained via an automatic mechanism involving priming at multiple levels of linguistic representation and, in addition, percolation of activation across these levels. Furthermore, alignment of situation models is achieved on the basis of such lower-level alignment of representations. While the model also allows for alignment via explicit reasoning and modeling of the partner's mental states and mental model updating, it places a particular emphasis on these low-level mechanisms. Alternative accounts of dialogue behavior question the explanatory power of automatic priming in dialogic coordination and underline the role of (explicit) modeling of partners and their mental states of representation. Thus, common conversational ground is the outcome of a joint effort on behalf of interlocutors who attend to the degree to which information is mutually shared [13]. Research on dialogue has addressed how speakers deal with variability and ambiguity in order to achieve alignment of situation models.

When dialogue partners refer to the same scene, they select a frame of reference or a perspective for the description. Thus, in dialogue, perspective use and perspective switching are part of the overall process of coordination. Does choice of perspective depend then on the previous use or preference for a certain perspective shown by one's dialogue partner, i.e., do speakers align in their choices of a spatial descriptive schema? If so, to what extent can this influence be modulated by the degree of consistency of partners' choices? Furthermore, how flexible is this process of coordination and perspective choice? Does the first 'conceptual pact' [14] one strikes implicitly with one's partner remain dominant throughout an interaction, and if your partner switches perspective, are you more likely to adhere to their initial perspective choice, or to switch along, and re-align spatial perspectives? Additionally, we ask, do these choices depend on the individual spatial ability of speakers? Would lower spatial ability be associated with more inconsistent use of perspective, or with a systematic preference for one kind of perspective over another? Alternatively, one might expect that an individual's higher spatial ability would make it possible for them to be more flexible, switch, and align, and re-align, and generally 'juggle' with perspective choices more easily than lower ability speakers would do.

In the studies presented here, there were two clearly possible perspectives on the scene and route to be described: survey (gaze) and route perspective. Route perspective is by far the more natural way to describe routes whereas survey perspective is more typical of location descriptions. In order to enhance the probability of use of the survey (gaze) perspective and to bring the two more into balance, the maps to be described were positioned vertically, which also corresponds to viewing maps on a screen.

In the first experiment, we ask first whether speakers align choices of spatial perspective when their partner follows one perspective consistently in a short sequence of descriptions (four maps with routes). We also ask whether spatial perspective alignment continues even when the partner switches perspectives and offers a subsequent series of descriptions in an alternative perspective and to what extent such behavior may be associated with different levels of spatial ability.

2 Experiment 1

As stated above, this experiment was designed to examine two related questions. First, whether speakers align on spatial perspective, and second, if they continue to align with their partners even when their partners switch perspective half-way through the experimental session, i.e., between an early and a later experimental block. If speakers rely on an enduring model of partner preferences formed upon first encounter, that is, built on the basis of their experience during the early block, then perspective switch by the confederate should not reverse speakers' choices in accordance with the new spatial perspective bias exhibited in the later block. If speakers are sensitive not only to initial partner preferences but they also update their model of their partner (after the switch), then they should also show a tendency to switch perspective in a similar way. A third possibility also exists—the fact that their partners have used both route and survey perspectives and that they switched between them may reduce speakers' preferences for either perspective and lead them to choose between perspectives more or less randomly.

2.1 Method

The design of the experiment included prime perspective (route vs. survey), order of prime experimental block (route-then-survey vs. survey-then-route) and spatial ability level (high vs. low in a median cut) as independent variables and mean percent choice of route perspective on each experimental block as the dependent variable.

Participants. 24 participants (3 male) took part in the experiment. They were university students with a mean age of 21.08 years (range 19 – 31) who received course credit or were paid for their participation. All were native German speakers.

Stimuli. Thirty-two simplified map drawings were used in the study. Six different maps were created and a total of 16 different routes. Stimuli were pseudo-randomized with the following constraints: no two consecutive maps (configurations of building blocks and streets) were the same, and neither the start nor the end points of the routes on consecutive maps were the same. Routes were pre-drawn on the maps so as to exclude a route planning component in the task and focus exclusively on the choice of spatial perspective (see Fig.1 for an example). There were 16 experimental trials (8 prime-target pairs) and 16 fillers. The maps and routes on the experimental prime-target trials were designed so as to be compatible with both route perspective and survey perspective descriptions. Confederates' descriptions of routes were in either route perspective or survey perspective. Filler maps and routes were drawn in such a way as to minimize the use of spatial perspective, for example, involving a circular trajectory. Furthermore, confederates' scripted descriptions on these trials did not contain any indication of spatial perspective.

Each experimental prime-target pair was preceded by two filler items. There were two blocks of experimental pairs, an early and a later one. The spatial perspective of the confederate primes was consistent within each block and was in either route or survey perspective. In addition, confederates' scripted descriptions on the two blocks differed in spatial perspective, i.e., the confederate switched perspectives between the early and the later block of trials.

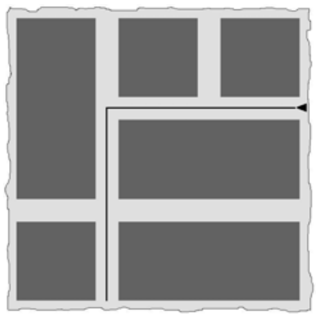


Fig. 1. Example of a map and its pre-drawn route. The triangle indicates the position and orientation at the start.

Procedure. Each participant was seated across a desk from the confederate with a visual barrier placed between them. A stack of cards with identical maps and routes drawn were placed in front of each of them on a vertical stand. In addition, the confederate used a list of pre-scripted descriptions matching their cards in either route or survey perspective. The scripted responses of the confederate were not visible. We took special care to minimize possible suspicions on behalf of the participants that their partner in the experiment may not be a naïve participant such as they were, including greeting routines, familiarization procedures, instructions, etc. Confederates were student assistants of the same age and population who were trained to act naïve. Participants and confederates took turns in describing the routes on these cards. A red and a green dot marked on the back of each card were used to indicate whose turn it was to speak. Confederates were the first to speak, thus ensuring that their utterances (primes) precede those of the participants on target trials. Participants were instructed to monitor the descriptions of their partner for accuracy and to offer a correction whenever they noted an incorrect description. Three deliberate errors were built into the script on filler items. This instruction ensured that participants were attending to their partners' descriptions. At the end of the experimental session, participants filled out a questionnaire which included questions asking what they thought the experiment was about and what they thought about their partner's behavior. As nobody indicated any suspicions that their partner may not have been a naïve participant such as they were, the data of all participants were accepted for analysis.

This procedural setup is a close replica of the procedure in [15] which studied the effects of participant role on syntactic alignment. Testing was conducted in the German language.

At the end of the experimental session, we also collected participants' responses to eleven 7-point scales of self-assessment of spatial ability, including the production and comprehension of route description, distance judgment, sense of orientation, use of cardinal reference points, map reading, etc. from the Freiburg Version of the Santa Barbara Sense of Direction Scale[1].

2.2 Results

The pre-analysis procedure was identical for the data in Experiment 1 and Experiment 2 and will be described here jointly. Participants' responses were classified according to the spatial perspective used as belonging to one of three categories: route perspective, survey perspective, or mixed perspective. Experimental prime-target pairs on which the confederate made a mistake (1.99%) were excluded from the analysis, as well as those where the participant offered a correction to their partner's description of an experimental item (20.39%). Route perspective was the preferred default option so in the majority of these cases participants offered a 'correction' of the confederate's survey perspective description into a route perspective one. The following are examples of participant responses to the map and route in Fig.1 coded as route perspective (a), survey perspective (b), and mixed perspective (c) in their original German and in translation:

(a) *hier gehst du geradeaus und biegst dann links ab*
E. here you go straight and then turn left

(b) *hier gehst du erst nach äh links und dann nach unten*
E. here you first go uhm left and then down

(c) *hier geht's geradeaus und dann nach unten*
E. here one goes straight and then down

The data for each participant for each block (early and late) were converted into mean percent use of route perspective.

Participants were classified into two groups of high and low ability based on the self-report measure which was the average score on the eleven items of the Freiburg Santa Barbara Sense of Direction scale. Twelve participants formed the high ability group, M = 4.45, SD = .39, range 3.91 – 5.27. Another twelve formed the low ability group. M = 3.15, SD = .41, range 2.55 – 3.82. A two-tailed t-test for independent samples confirmed that the two groups differed significantly in terms of their self-assessment score on this measure, $t(22) = 8.01$, $p < .001$.

[1] Due to technical error, data on the last four of the fifteen items were not collected.

A mixed between-within participants analysis of variance was conducted to assess the impact of prime perspective (route vs. survey), order of prime perspective (route on the early block and survey on the later block vs. survey on the early block and survey on the later block) and individual spatial ability (high vs. low ability on the sense of direction scale) on the mean percent use of route perspective. There was a main effect of prime perspective, Wilks Lambda = .46, $F(1,20) = 23.34$, $p < .001$, partial eta squared = .54, and, more importantly, a significant three-way interaction of prime perspective, order of prime, and spatial ability, Wilks Lambda = .72, $F(1,20) = 7.89$, $p = .011$, partial eta squared = .28.

To unpack the interaction, separate analyses were conducted for each of the two spatial ability groups (high and low ability). A repeated measures analysis of percent use of route perspective with order of prime as a between participant variable and prime perspective as a within participant variable in the *high* ability group revealed no interaction between these two factors, Wilks Lambda = . 93, $F(1,10) = .81$, p = . 39, and an effect of prime, Wilks Lambda = .42, $F(1,10) = .14.02$, $p < .01$, partial eta squared = .58. There was no difference between the two orders of prime perspective, $F(1,10) = 2.35$, $p = .16$. A repeated measures analysis of variance on the mean percent use of route perspective with the same between and within participant variables for the *low* ability group resulted in a significant two-way interaction of prime perspective and order of prime, Wilks Lambda = .55, $F(1,10) = 8.34$, $p < .05$, partial eta squared = .45.

Generally, participants' use of spatial perspective was guided by the perspective they heard their partner use (the prime perspective) – the mean percent use of route (vs. survey and mixed) perspective was much higher overall in the route prime condition than the survey prime condition. Furthermore, the high ability participants exhibited a general tendency to adapt to their partner's choices in the direction of priming on the later block (after the switch), decreasing their use of route perspective (from 83.33% to 47.17%) when their partner switched from route to survey perspective, and increasing their use of route perspective (from 72.33% to 94.50%) when their partner switched from survey to route perspective. The low ability participants, however, only switched along when their partner changed their perspective choices from survey to route perspective (from 37.50% to 95.83%) and not vice versa (94.50% vs. 91.67%). They persisted in using the route perspective in their descriptions on the later block even after their partner had shifted towards survey perspective use (see Table 1).

2.3 Discussion

Experiment 1 showed that speakers do align spatial perspectives with their partners. Those who heard their partner use a survey perspective consistently on the early block of four consecutive experimental trials were less likely to adhere to the otherwise preferred default of route perspective and used instead survey perspective themselves or a mix of the two perspectives in their descriptions. This effect was especially pronounced in the low ability group where the difference in the mean percent use of route perspective in the two prime conditions on the early block was 57%. In comparison, there was less variation in the choice of perspective as a function of prime perspective for the participants from the high ability group (only 11% difference). However, generally speaking, the priming effect was not confined to the early experimental block

Table 1. Mean percent use (standard deviations in brackets) of route perspective before and after perspective switch in Experiment 1 for high and low spatial ability (sense of direction, SOD) for the two orders of prime (route-survey and survey-route)

	High ability (SOD)		Low ability (SOD)	
	Early block pre-switch	Later block after switch	Early block before switch	Later block after switch
Route – Survey order	83.33 (20.41)	47.17 (32.78)	94.50 (13.47)	91.67 (20.41)
Survey – Route order	72.33 (27.70)	94.50 (13.47)	37.50 (49.37)	95.83 (10.21)

only. Perspective priming also occurred on the later block of four target trials as well. This is particularly striking in view of the nature of the second (later) experimental pairs. During those trials, the confederate used the alternative perspective to the one he or she used on the early block, thus displaying a switch from survey to route perspective or vice versa. In this sense, although on each set of four consecutive trials the confederate had made consistent perspective 'choices' in their descriptions, across the two experimental blocks their behavior appeared inconsistent, and yet, participants had the same tendency to align with their partners later as well as earlier during the experimental session. This is notable for two reasons. First, it shows that spatial perspective is used flexibly, and that speakers make use of the possibility to switch perspective with relative ease. Second, it also shows that speakers were not entrained on the first perspective only that they heard their partner use but that they updated. In this sense, this experiment has provided evidence for speakers' sensitivity to their partners' changes in behavior and preference for a representation scheme.

Perspective priming, however, occurred in rather different patterns in the two (high and low) ability groups. First, as noted above, unlike high ability participants, low ability participants showed a very large priming effect on the early block with a 57% difference in the mean percent use of route perspective between the two prime conditions. The patterns diverged further in an interesting way on the later experimental block. While the perspective prime produced by the confederate resulted in a 47% gap in perspective choices for the high ability participants, there was hardly any difference across the two prime conditions for the low ability group. One possible interpretation of these results would be that the low ability participants did not align in spatial perspective with their partner after their partner switched perspective preference between the early and the later block. A closer look at the pattern of results separately for the two orders of prime, however, reveals a more complex picture. The low ability participants who had been exposed to priming by the *route* perspective on the early block did not change their preference, indeed, and continued to use mostly the same route perspective even when their partner switched clearly to survey perspective. The low ability participants who received perspective priming in the *survey* perspective on the early block, on the other hand, and who as a group had shown an interactive alignment effect on that early block, did show such an effect after their partner switched from survey perspective on the early block to route perspective use on the later block. That is, absence of interactive alignment as such is not sufficient to explain parsimoniously the pattern of perspective use in the low ability group after the switch. Instead, two mechanisms appear to have a combined influence on their choices, alignment of spatial perspective with one's partner being one of them when we compare choices on the early and later blocks. The second source of influence is the general preference for one of the perspectives in this study, namely, the route, or path-embedded perspective. Once established as a valid and consistently used choice by the partner in an interaction, it appeared to be much more preferred by the low ability participants even after the switch. Route perspective dominated participants' descriptions generally, and when they made corrections to their partner's (confederate) descriptions, they were almost exclusively on trials where the confederate used survey perspective and participants 'corrected' their descriptions by suggesting that the alternative route perspective should be used instead.

One might be tempted to argue that in this sense our low ability participants have shown a consistency effect in that they consistently adhered to one and the same perspective throughout the experiment, i.e., in the face of both route and survey perspective primes. This interpretation, however, can only be limited to one of the two orders of prime, route-then-survey, as low ability participants did not stick to survey perspective throughout the study in the survey-then-route order of primes. That is, intra-individual consistency of perspective choice was salient for the route perspective but not for the survey perspective here. We interpret this as a markedness, or default bias, effect. When one of two choices is the default, preferred (unmarked) choice, then switching away from this option after it has been used previously may be costly and require more motivation whereas switching from the dispreferred (cognitively marked) option to the preferred default option would be easier to accomplish and more likely. In this study, route perspective was clearly the dominant, unmarked option and survey perspective was the dispreferred, marked one. All in all, what this amounts to is that for the low ability participants switching along with their partner in the route-then-survey order may have been more difficult and/or seen as less necessary and therefore not typical of their behavior.

On the other hand, if consistency were to exert a stronger influence on spatial choice than interactive alignment, then one would not expect to find much perspective priming after their partner's switch in the high ability group who could stick to their partner's original perspective choice on the early block even when that was in the dispreferred survey perspective, given that they may be less adversely affected by having to use the marked, dispreferred perspective. However, the data support an explanation where consistency of perspective use is a less powerful mechanism in an interactive situation such as the confederate setup in our study than in the monologic settings used in much of the literature so far [6, 10].

In summary, on the early block, before the switch, low ability participants showed a large perspective priming effect by adapting considerably to the way perspective was used by their partner but after their partner's switch this form of interpersonal adaptation was modulated by the degree to which the two perspectives (route vs. survey/mixed) were dispreferred in the task in that they adapted easily to a switch away from the dispreferred to the preferred perspective (from survey to route) but not the other way round. A different pattern emerged in the high ability group who showed a general priming effect independent of the order of experimental blocks and spatial perspectives, i.e., a general ability and readiness to adapt to their partner and to switch perspectives.

3 Experiment 2

The results of the first experiment provided evidence for speakers' alignment with their partner at the conceptual level of spatial perspective both before and after their partner switched from an initial perspective (route or survey) to its alternative. However, within each of the two experimental blocks, confederates adhered consistently to one perspective only. Thus, although they switched perspective on the later block, prime perspective also remained constant for all four experimental pairs within that block. It is not clear, however, whether speakers' alignment on spatial perspective may have been influenced by this high degree of consistency within an experimental block. The second experiment set out to test whether speakers would also show conceptual alignment of spatial perspective with their partner even if the partner showed high inconsistency and switched perspective all the time, that is, between trials rather than between experimental blocks (as in Experiment 1). Constantly switching spatial perspective may make the confederate's choices appear more random and may thus lead participants to adopt a generally 'random' choice approach themselves. To distinguish between this possible outcome and systematic alignment with one's partner even in the face of the partner's inconsistency, we conducted the second experiment and ran an analysis of speakers' choices as a function of the immediately preceding prime for each target item.

3.1 Method

The design of the second experiment was simpler than the first one. It included prime perspective (route vs. survey) as an independent variable and mean percent choice of route perspective as the dependent variable. However, prime perspective in this case was inconsistent, i.e., constantly alternating between trials. In addition, we also examined the impact of individual spatial ability on participants' behavior here as well.

Participants. 19 participants (3 male) took part in the experiment. They were university students with a mean age of 21.32 years (range 19 – 28) who received course credit or were paid for their participation. All were native German speakers.

Stimuli. The same visual stimuli were used as in Experiment 1. However, in this second experiment, the confederate switched perspective on each trial. The first description they gave was in route perspective in one of the experimental lists and in survey perspective in the other list. Thus, the perspective of the confederate primes was inconsistent across trials within each of the experimental blocks (both early and late).

Procedure. The procedure was identical to the one used in Experiment 1.

3.2 Results

As in the first experiment, participants were classified into two groups of high and low ability based on the self-report measure which was the average score on eleven items of the Freiburg Santa Barbara Sense of Direction scale. Ten participants formed the high ability group, $M = 4.62$, $SD = .40$, range 3.91 – 5.36. Another nine formed the low ability group, $M = 3.48$, $SD = .34$, range 2.82 – 3.82. A two-tailed t-test for independent samples confirmed that the two groups differed significantly in terms of their self-assessment score on this measure, $t(18) = 6.58$, $p < .001$.

Participants' responses were classified according to the spatial perspective used as in Experiment 1. The data for each prime condition (route and survey perspective prime) for each participant were converted into mean percent use of route perspective. Only sixteen participants' data were included in the analysis as there were no valid data for the remaining three in the survey prime condition.

A mixed between-within participants' analysis of variance was conducted to assess the impact of prime perspective (route vs. survey) and individual spatial ability (high vs. low) on the mean percent use of route perspective. There was no interaction between the two factors, Wilks Lambda = 1, $F(1,14) < .01$, $p = .98$, and although the influence of prime perspective on the mean percent use of route perspective by participants was in the expected direction, the effect did not reach significance, Wilks Lambda = .81, $F(1,14) = 3.19$, $p = .10$, partial eta squared = .19 (see Table 2). In the survey prime condition, participants in the high ability group produced descriptions in the route perspective on average 66.67% of the time in comparison with their descriptions in the route prime condition where they used the route perspective 80.56% of the time (see Table 2). Similarly, participants in the low ability group produced fewer descriptions in the route perspective when their partner used the survey perspective (82.14%) than when their partner used the route perspective (96.43%); although this was not a statistically significant effect.

Table 2. Mean percent use (standard deviations in brackets) of route perspective in Experiment 2 for high and low spatial ability participants (sense of direction, SOD) for the two prime conditions (route and survey perspective)

	High ability (SOD)	Low ability (SOD)
Route prime	80.56 (39.09)	96.43 (9.45)
Survey prime	66.67 (41.46)	82.14 (27.82)

It is worth reiterating here that the data of only sixteen of the participants (nine with high ability and seven with low ability) in the experiment were accepted for this analysis because of insufficient data in the survey prime condition for the remaining three participants. Their responses on these trials could not be used as valid data mostly due to the numerous corrections offered by these participants to their partner's choice of perspective.

3.3 Discussion

The general finding of the first experiment was that speakers aligned in spatial perspective with their conversation partners, although this effect was mediated by individual spatial ability. However, one good reason for the emergence and persistence of alignment may have been that the behavior of the speakers' partners within experimental blocks remained consistent. If spatial perspective is used inconsistently by one's partner, this kind of information may be less reliable and relevant for speakers to be primed by. Experiment 2 put this possibility to the test. Here confederates' prescripted descriptions switched between the two perspectives constantly, i.e., if their first, third, fifth, etc. utterances were in a route perspective, then their second, fourth, sixth descriptions were in a survey perspective, and vice versa. The analysis of the data revealed that participants did not align interactively even though on average they used more route perspective descriptions after they heard their partner use route perspective on the immediately preceding trial than if they heard their partner use a survey perspective. This difference, however, did not yield a statistically significant effect. Furthermore, individual spatial ability did not make a difference, either. Therefore, in a comparison of the findings of experiment 1 and experiment 2, we can conclude that spatial perspective alignment is constrained by the relative degree of consistency of speakers' descriptions, although not necessarily by perspective switching. That is, speakers may still be primed by the spatial perspective they heard even when their partner made a switch after having produced a number of consistent descriptions, as in experiment 1, but not when the partner uses perspective inconsistently, switching to and fro all the time.

However, this interpretation can only be offered with a proviso. As described earlier, items where the participant objected to the description used by their partner were not included in the data set for analysis as priming could not be tested because of an interruption of the direct prime-target sequence. Nevertheless, the important finding from Experiment 2 was that speakers did not align in spatial perspective when their partners exhibited a highly inconsistent descriptive behavior by constantly switching between the two perspective schemas. Such inconsistency by the partner may have led participants to view either perspective as equally suitable and then adopt a less effortful strategy.

4 Conclusion

In two experiments, we sought to establish whether speakers align with their partners on the spatial perspective used (path-embedded vs. environment-external) in simple route descriptions of a series of map drawings. The data in the first study generally confirmed this hypothesis while in the second study it was not supported reliably. The critical difference between the two experimental designs was the level of consistency of partners' use of spatial perspective. When partners used perspective consistently in a series of descriptions, speakers were more likely to align their perspective choices with them. When this was not the case, as in experiment 2 where their partners used spatial perspective inconsistently, speakers were not influenced significantly by their choices. However, if partners switched perspective only once after having been consistent for a block of four trials, speakers still generally aligned with them. Thus, switching itself did not lead to an absence of perspective alignment, inconsistent use did.

To what extent can we attribute these results to automatic priming as in the interactive alignment framework or to a conscious strategy and mental modeling on behalf of speakers? Although it may be difficult to draw a clear line between the two interpretations, the fact that participants re-aligned after their partners switched perspective suggests that automatic priming from repeated exposure to a certain perspective can be overridden by more powerful mechanisms such as precedent violation here, or breaking a 'conceptual pact' established earlier [14]. Unless we postulate individual variability in susceptibility to priming, the interaction of individual spatial ability with spatial perspective alignment as seen in experiment 1 also paints a picture more complex than mechanistic priming as the only underlying mechanism.

The importance of individual differences in spatial cognition has long been in the focus of research attention. Here we also examined such differences. The sample of participants in the two experiments was not balanced enough to allow us to draw clear conclusions on gender differences in speakers' alignment behavior. [2] However, we have seen that individual spatial ability played an important role in the process of spatial perspective alignment. Low ability participants were more likely to be influenced by the initial choice of perspective of their partner on the early block before the switch but less so on the later block, in particular they re-aligned with their partner and switched along only if the switch was from survey to route perspective but not the other way around. Once established as a viable option by their partners' descriptions before the switch, route perspective dominated low ability speakers' choices until the end. In contrast, high ability participants were influenced by priming in both perspectives after the switch and re-aligned systematically throughout the study and independently of the perspective used. Thus, low ability participants were sensitive to partners' initial choices but showed less ability and/or readiness to re-align perspectives after their partners switched. Although we may be tempted to interpret the diverging patterns of low and high ability participants in terms of cognitive load which may have affected participants in the two groups to a different extent, presumably producing more disadvantages for the low ability group, this may not be the only explanation. As Tversky, Lee, & Mainwaring [1] have argued, the cognitive costs of switching perspective may be balanced by the cognitive costs of retaining the same perspective, especially in situations where switching perspective may be more effective in communication than not switching perspective. More work needs to be carried out before we can draw more informed conclusions on the relative benefits of consistency vs. flexibility in perspective use, in general, and in dialogic settings, in particular, where the issue is related to interpersonal adaptation.

Further research into spatial perspective alignment will help solve more mysteries. A memory task experiment [16] has revealed spatial perspective priming. A comparison of the two studies indicates common underlying mechanisms that need to be explored further.

---

[2] There were only 6 male participants in the two experiments (3 in experiment 1, and 3 in experiment 2). Interestingly enough, all six used the route perspective throughout their descriptions despite the fact that they were in different experimental prime conditions.

Acknowledgments

Thanks go to Christoph Hölscher of the University of Freiburg for the use of the Freiburg Version of the Santa Barbara Sense of Direction Scale, and to the research assistants, confederates, and the research team of I5-DiaSpace. The financial assistance of the German Science Fund (DFG) is acknowledged for funding provided to SFB/TR8 Spatial Cognition at the University of Bremen where this research was conducted. The comments and suggestions of the anonymous reviewers of an earlier version of this manuscript are gratefully acknowledged.

References

1. Tversky, B., Lee, P.U., Mainwaring, S.: Why Speakers Mix Perspectives. Journal of Spatial Cognition and Computation 1, 399–412 (1999)
2. Kriz, S., Hegarty, M.: Spatial Perspective in Spoken Descriptions of Real World Environments at Different Scales. In: Proceedings of XXVII Annual Meeting of the Cognitive Science Society, Stresa, Italy (2005)
3. Taylor, H.A., Naylor, S.J., Chechile, N.A.: Goal-specific Influences on the Representation of Spatial Perspective. Memory & Cognition 27(2), 309–319 (1999)
4. Taylor, H.A., Tversky, B.: Perspective in Spatial Descriptions. Journal of Memory and Language 35, 371–391 (1996)
5. Bugmann, D., Coventry, K.R., Newstead, S.E.: Contextual Cues and the Retrieval of Information from Cognitive Maps. Memory & Cognition 35(3), 381–392 (2007)
6. Vorwerg, C.: Consistency in Successive Spatial Utterances. In: Coventry, K.R., Tenbrink, T., Bateman, J. (eds.) Spatial Language and Dialogue (2009)
7. Schober, M.F.: Spatial Perspective-taking in Conversation. Cognition 47, 1–24 (1993)
8. Striegnitz, K., Tepper, P., Lovett, A., Cassel, J.: Knowledge Representation for Generating Locating Gestures in Route Directions. In: Coventry, K.R., Tenbrink, T., Bateman, J. (eds.) Spatial Language and Dialogue (2009)
9. Watson, M.E., Pickering, M.J., Branigan, H.P.: Alignment of Reference Frames in Dialogue. In: Proceedings of the 26th Cognitive Science Meeting. Erlbaum, Mahwah (2004)
10. Levelt, W.J.M.: Cognitive Styles in the Use of Spatial Direction Terms. In: Jarvella, R.J., Klein, W. (eds.) Speech, Place, and Action. John Wiley & Sons, Ltd., Chichester (1982)
11. Schober, M.: Spatial Dialogue between Partners with Mismatched Abilities. In: Coventry, K.R., Tenbrink, T., Bateman, J. (eds.) Spatial Language and Dialogue (2009)
12. Pickering, M., Garrod, S.: The Interactive Alignment Model. Behavioral and Brain Sciences 27(2), 169–189 (2004)
13. Clark, H.H.: Using Language. Cambridge University Press, New York (1996)
14. Brennan, S.E., Clark, H.H.: Conceptual Pacts and Lexical Choice in Conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition 22(6), 1482–1493 (1996)
15. Branigan, H.P., Pickering, M.J., McLean, J.F., Cleland, A.A.: Syntactic Alignment and Participant Role in Dialogue. Cognition 104, 163–197 (2007)
16. Andonova, E., Coventry, K.: Alignment and Priming of Spatial Perspective. In: Edlund, J., Gustafson, J., Hjalmarsson, A., Skantze, G. (eds.) Proceedings of Diaholmia, Workshop on the Semantics and Pragmatics of Dialogue. KTH, Stockholm (2009)

The Role of Grammatical Aspect in the Dynamics of Spatial Descriptions

Sarah Anderson[1,*], Teenie Matlock[2], and Michael Spivey[2]

[1] Department of Psychology, Cornell University, Uris Hall, Ithaca, New York 14853 USA
sec57@cornell.edu
[2] Cognitive and Information Sciences, University of California, Merced, Merced, California 95344
{tmatlock,spivey}@ucmerced.edu

Abstract. What role does grammatical aspect play in the time course of understanding spatial language, in particular motion events? Although processing differences between past progressive (was walking) and simple past (walked) aspect suggest differences in prominence of certain semantic properties, details about the temporal dynamics of aspect processing have been largely ignored. The current work uses mouse-tracking [1] to explore spatial differences in motor output response to contextual descriptions and aspectual forms. Participants heard descriptions of terrain (difficult or easy) and motion events described with either the past progressive or simple past aspectual form while placing a character into a scene to match this description. Overall, terrain descriptions modulated responses to past progressive more than to simple past in the region of the screen corresponding to the path. These results, which suggest that perceptual simulation plays a role in the interpretation of grammatical form, provide new insights into the understanding of spatial descriptions that include motion.

Keywords: Spatial language, Motion verbs, Event understanding, Mouse-tracking, embodied cognition.

1 Introduction

Language and space are intimately connected. All languages have a system for specifying basic spatial relations, including where things are relative to other things and where they are relative to time. Some of this happens by concatenating lexical items to indicate where an object is at any given moment. For instance, the expression "at the end" in "John is at the end of his driveway", tells us that John is standing, sitting or otherwise positioned in the end region of his driveway at the time of speaking. All languages have a way to express movement, including movement to and away from landmarks, for instance, "John went to the end of this driveway" or "John is coming from the end of his driveway". In these cases, an agent (John) moves from one position to another in space and time. Spatial expressions like these are so ubiquitous in language that we rarely give them a second thought when we are describing or listening to descriptions of physical space. And despite a huge and well-informed literature

* Corresponding author.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 139–151, 2010.

on how spatial descriptions are used in linguistics and psychology, surprisingly little is known about how grammar influences the way they are processed in real-time. In particular, what role does grammatical aspect play in the interpretation of past events, especially motion events? When is it important to highlight the ongoing nature of an event versus its completion, and what consequences does this often implicit distinction have for understanding where agents and objects are at any given movement? This research explores these questions using mouse-tracking, a robust dense-sampling method designed to track the fine-grained details of spatial processing in a number of cognitive domains [1]. This method allows us to pinpoint where things are at any given moment relative to linguistic input and physical space.

Grammatical aspect provides temporal information about the unfolding of events. In this way it provides the speaker and listener implicit, detailed temporal information about a reported event, for instance, whether it was completed and whether it was a short or long [2, 3]. Such temporal information can certainly influence the way a sentence is understood. Take, for example, the following sentences: “David walked to the university,” and “David was walking to the university.” Both sentences describe an event that occurred in the past, but they use different aspectual forms. The first uses the simple past form of the verb “ran,” and emphasizes the completion of the action. The second uses the past progressive form, and emphasizes the ongoing nature of the action. The forms verb+ed and was verb+ing are often used to describe the same situation in English. (In some languages, this perfective versus imperfective distinction is pronounced and highly differentiated. In English, the distinction is much more diffuse. See Croft [4] and Frawley, [3].) However, just how they differ in terms of processing is understudied and many questions remain about how they influence spatial reasoning.

In recent work, we began to explore how aspect influences spatial reasoning in a series of simple offline studies [5]. In each study, people were shown a picture of a path that extended from the foreground of the scene into the background of the scene where it terminated at a destination landmark (e.g., a university, mountain range). Below the picture was a sentence that related to the scene shown in the picture. The sentence included a moving agent and a motion verb described with either the simple past or past progressive aspectual form, for instance, “This morning David walked to the university” (simple past) or “This morning David was walking to the university” (past progressive). On the path were 10 unevenly spaced identical silhouette characters who appeared to be heading to the destination landmark (e.g., figure with leg extended forward and arms bent as if in motion and facing the university). Participants were instructed to “circle the man that the sentence is most naturally referring to.” The results showed that participants were more likely to circle a character in the middle region of the path with past progressive sentences (for instance, was walking), and more likely to circle a character in the latter region of the path in response to simple past sentences (for instance, walked). These results indicated that when participants read simple past sentences, they focused on the end of the path, or the location of the completed action in the scene. In contrast, when they read past progressive sentences, they focused on the middle section of the path, where the ongoing action would be taking place. In these and related offline studies, we showed that different aspectual forms have clear consequences for spatial reasoning, in particular, the conceptualization of motion events.

The findings of our offline studies resonated with reaction time studies conducted by Madden and Zwaan [6]. In analyzing the on-line processing of simple past and past progressive event descriptions in a sentence-picture verification task, they found that people were generally quicker to respond to pictures showing a completed action after they reading a simple past sentence (e.g., “The car sped through the intersection”) versus a past progressive sentence (e.g., “The car was speeding through the intersection”). However, no such latency differences were found when participants read past progressive sentences and saw pictures of intermediate action. In brief, their studies showed that the simple past draws attention to the end state of the action and the past progressive draws attention to a wide and varied range of intermediate stages of an action. Although their work provided insights into the processing of aspect and its impact on spatial reasoning, many questions remained about the details of processing. (For related work on aspect and spatial representation, see 7, 8, and 9).

Such reaction time data have revealed valuable insights into the processing of aspect. However, as suggested by the work of Madden and Zwaan [6], they are somewhat limited when investigating details of representation (especially with the past progressive, on-going form). In addition to considering the findings of the offline and reaction time experiments mentioned above, it is necessary to consider relevant research on real-time cognitive processing in the dynamics of the response. To gain new insights into the role of aspect in event understanding, we consider relevant work on real-time cognitive processing in response dynamics. Factors that can affect response latencies have been shown to also have the capacity to influence later facets of response dynamics. For instance, Abrams and Balota [10] demonstrated that word frequency is capable of influencing not only response latencies but also response kinematics after a response has been initiated, suggesting ongoing language processes may co-exist with ongoing motor processes (see also 11, 12 and 13). Their work makes a strong case for examining the dynamic variables of motor movements initiated in response to a stimulus.

To better understand the potential differences in the on-line processing of different aspectual forms, here we use the methodology of computer-mouse tracking. Monitoring the streaming x- and y-coordinates of goal-directed mouse movements in response to spoken language is a useful indicator of underlying cognitive processes. In contrast to ballistic saccades, arm movements allow for a continuous, smooth motor output within a single trial to complement eye-tracking research. Spivey, Grosjean, and Knoblich [1] demonstrated that these mouse movements can be used to index the continuous activation of lexical alternatives. By recording the x,y coordinates of the mouse as it moved with the goal-directed hand motion to click on the appropriate object, competition between the partially activate lexical representations was revealed in the shape and curvature of the hand-movement trajectories.

Further, Dale, Kehoe, and Spivey [14] employed this mouse-tracking methodology to explore the underlying processing of categorization. In a series of four experiments, participants used the mouse to click on one of two categories (e.g., “mammal” or “fish”) to categorize either a typical exemplar (“cat”) or an atypical exemplar (“whale”), while the computer-mouse movements were recorded. The results showed spatial differences in the average trajectories of the two conditions, with the atypical exemplars’ average trajectory diverging away from the typical exemplars’ average trajectory towards the competing category response button. Additionally, the movement durations for each condition were significantly different for the two conditions, with the atypical trajectories having longer overall movement durations than the typical trajectories. These results reveal nonlinear time course effects in the process of categorization, and significant attraction towards the competing category name in the atypical exemplar trajectories. Moreover, they provide evidence to complement reaction time data by examining the overall movement durations of the two types of trajectories.

Some of our own work indicates that mouse-tracking is useful and informative for exploring research questions on the on-line processing of grammatical aspect [15]. In one experiment, participants listened to sentences like, "Tom jogged to the woods and then stretched when he got there," or "Tom was jogging to the woods and then stretched when he got there." While participants heard these sentences, they saw scenes consisting of a path curving upwards from left to right, and terminating at the destination described in the sentence. A character was located to the right of the beginning of the path and under the destination, separated from the scene by a black box framing the destination and path. Similar to our earlier offline results, participants dropped the character closer to the center of the path with past progressive sentences and closer to the destination with simple past sentences. Further, the two aspectual forms elicited significantly different movement durations: Participants moved the character into the scene for a longer duration of time with past progressive sentences than when they heard simple past sentences. These drop location and movement duration results converge with and further inform earlier research, supporting that past progressive aspect focuses attention on the on-going nature of the action while simple past aspect focuses attention on the end state of that action, even during real time processing.

In the current experiment, we set out to extend these findings by investigating the way aspect may interact with terrain descriptions. Research has shown that context descriptions interact with fictive motion sentences to produce both differences in patterns of eye movements and in reaction times [16, 17]. Eye movement data have also elucidated the real-time comprehension of fictive motion sentences. Fictive motion sentences contain motion but no actual movement takes place. For example, "The road ran through the valley" contains ran, but the road does not actually run. Richardson and Matlock [17] had participants listen to a context sentence describing the terrain, and then a target sentence containing fictive motion. When participants first heard context sentences describing the terrain as difficult, inspection times and eye-movement scanning along the path were increased as opposed to when participants heard the terrain described as easy. These results support earlier reaction time results, in which participants read narratives describing a terrain, and then made decisions about whether a fictive motion sentence was related to the preceding context [16]. Participants were slower to respond to fictive motion sentences when they had read context sentences describing slow travel, long distances, or difficult terrain, than when the they had read contexts describing fast travel, short distances, or easy terrain. However, no such differences in context descriptions was found when the target sentence did not contain fictive motion (e.g, "The road is in the valley"). Taken together, the results provide evidence that fictive motion descriptions affect both reaction times and eye-movements by evoking mental representations of motion, and that this is then influenced by contextual constraints on that motion. The eye-movement data allow for a closer look at the way the constructed mental model and the linguistic description are coordinated.

However, the impact and coordination of such descriptions and grammatical aspect has not been explored. Here we use the mouse-tracking methodology to investigate how different aspectual forms interact with similar context descriptions. Participants heard two sentences. The first provided a contextual description of the path and the second manipulated grammatical aspect. For example, on target trials participants heard a context sentence describing the path as either difficult (i.e., "The road to the university was rocky and bumpy") or easy (i.e., "The road to the university was level and clear"), before a simple past target sentence (i.e., "David walked to the university where he sat in class") or a past progressive verb target sentence (i.e., "David was walking to the university where he sat in class"). At the same time participants heard these sentences, participants saw scenes like Figure 1 below.

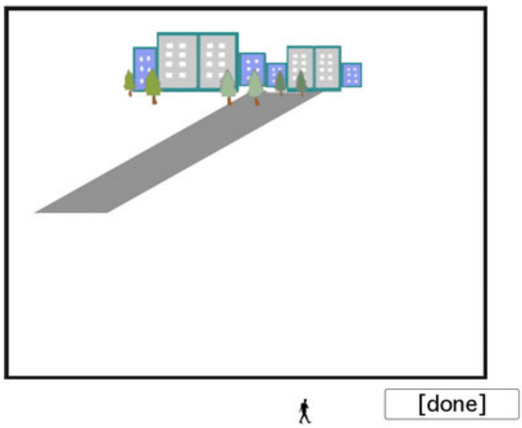


Fig. 1. In our experiment, the auditory stimuli (Table 1) were accompanied by a visual scene, like this example

This scene contained a diagonal path that originated halfway up the screen and extended from the extreme left to the top and center of the screen (corresponding to the destination in the sentence). The orientation of the path was changed to this short, diagonal path from the long, curvy path of earlier research [5, 15] to allow for more thorough and precise investigations of potential spatial and movement duration differences. A character was located to the right of the beginning of the path and under the destination. It was outside of the scene (i.e., separated by a black box that framed the destination and path).

We explored several hypotheses around the role of grammar in the processing of spatial descriptions and motion conceptualization. If past progressive sentences elicit more attention to the path, then the effect of context description was expected to be greater with past progressive sentences than when they contained simple past sentences. We predicted that context would modulate movement durations and spatial attraction to the path more in the past progressive sentences than in the simple past sentences. We also wanted to explore the influence of the visual scene's path on movement durations. The visual scene---with a path starting halfway up the screen---would enable us to examine if the trajectories produced in response to each aspectual form would reliably differ for the entire trajectory of the hand or only in the region of the screen corresponding to the path. If differences emerged across the entire trajectory, then the effect of grammatical aspect would appear to be more global, and to exert influence across the entire event description. However, if differences emerged only in the region of the screen corresponding to the path, then the effect of grammatical aspect would appear to be specific to the parts of the event it describes.

2 Method

2.1 Participants

A total of 64 undergraduates at Cornell University participated in the experiment for extra credit in psychology courses. All participants were right handed and native speakers of American English.

2.2 Materials

Twelve sentences were created from adapting the stimuli used in the offline studies of Matlock and colleagues [15]. They contained a range of motion verbs, including walk, job, run, and hike. A final clause that described an event at the destination was added to encouraging movement all the way to the destination. Similarly, two contexts for each stimulus were created. Hence, four versions of each of the 12 experimental items were created, as shown below: (1a) rough context description, simple past sentence, (1b) rough context description, past progressive sentence, (1c) smooth path description, simple past sentence, (1d), smooth path description, past progressive sentence.

1a) *The road to the university was rocky and bumpy.* David walked to the university where he sat in class.
1b) *The road to the university was rocky and bumpy.* David was walking to the university where he sat in class.
1c) *The road to the university was level and clear.* David walked to the university where he sat in class.
1d) *The road to the university was level and clear.* David was walking to the university where he sat in class.

Sentences were recorded using a Mac-based speech synthesizer program. Each of the 12 experimental items was spliced in order to produce both a past progressive and a simple past version, ensuring that the prosody of both of the targets was otherwise identical. Similarly, the context description was spliced onto the beginning of each of these target sentences. A pause of one second separated the offset of each context sentence from the onset of the target sentence. The experimental items were counterbalanced across four presentation lists. Each list contained three instances of each condition, so that all participants heard all twelve target sentences, but only heard one version of each.

Corresponding visual scenes were created for each target sentence pair. Each target visual scene consisted of a diagonal path starting halfway up and on the extreme left side of the screen. The path slanted to the right, terminating in the middle at the top of the screen. A character was located to the right of the beginning of the path and under the destination, separated from the scene by a black box framing the destination and path. See Figure 1 above. The only moveable item in the scene was the character, which subtended an average of 1.53 degrees of visual angle in width by 2.05 degrees in height. The destinations were an average of 11.22 degrees of visual angle in width by 4.09 degrees in height, and the path itself occupied a square of 8.42 degrees of visual angle in width by 6.11 degrees of visual angle in height. The character was located 14.25 degrees of visual angle from the destination. The stimuli were presented using Macromedia Director MX, and mouse movements were recorded at an average sampling rate of 40 Hz. The display resolution was set to 1024 x 768.

Additionally, to keep participants from developing strategies specific to the experimental sentences, 12 filler items were created. The fillers were of the same form as the target sentences: each contained a context description and either a past progressive or simple past sentence. These filler trials varied from the target trials such that the context description provided no information about the path (i.e., "The weather in the valley was warm and humid") and such that they described no movement along the path (i.e., "Janet swam in the pool and then dried in the sun,"). These filler items were accompanied by 12 filler scenes, created using a short path beginning on the right side of the screen and slanting to the top, center of the screen. Besides the direction of the path, each filler scene was quite similar to the target scenes, for instance, character outside of a scene that contained the path and destination mentioned in the filler sentence.

2.3 Procedure

Participants were asked to make themselves comfortable in front of the computer, and allowed to adjust the mouse and mouse-pad to a location that suited them. First, participants read instructions to place the character in the scene to make the scene match the sentences they heard. Upon signaling to the experimenter that they understood the task, they were next presented with two practice trials (similar in form to the filler trials), followed by the experimental task. At the onset of each trail, participants were presented with the entire visual scene. After a 500 ms preview, the sound file began. After the participant had moved the character (though not to any particular location), a "Done" button appeared in the bottom left corner of the screen. Participants clicked this button to move to the next trial. A blank screen with a button in the center labeled "Click here to go on" separated trials from each other. The entire experiment lasted approximately 20 minutes.

3 Results

Mouse movements were recorded during the grab-click, transferal, and drop-click of the character in the experimental trials. Prior to the analyses, the data were screened to remove extremely long trials. Movement durations 20 seconds or more were removed because they constituted an unusually long time for a mouse-movement. Using this criteria, only three trials (less than 0.4%) of trajectories, were excluded.

3.1 Drop Locations

Previous offline results revealed that participants chose a location closer to the middle of the path as the best representative of a past progressive sentence, while selecting a location closer to the destination as the best representative of a simple past sentence [5, 15]. By plotting the drop point (location along the path where each participant let go of the mouse to "drop" the character) in each of the four conditions, the current results demonstrate a similar trend. See Figure 2. There was not a significant interaction of terrain description and aspect (p's > .5). However, there was a main effect of aspect when comparing the average drop x-coordinate, $F(1,62) = 8.462$, $p < 0.005$, with the average drop x-coordinate being further left (closer to the path) when participants heard past progressive sentences ($M = 476.71$, $SD = 68.81$) than when they heard simple past sentences ($M = 494.82$, $SD = 61.74$). Similarly, there was a main effect of aspect when comparing the average drop y-coordinates, $F(1,62) = 6.048$, $p < 0.017$, with the average drop y-coordinate being lower (further from the destination) when participants heard past progressive sentences ($M = 219.04$, $SD = 37.02$) than when they heard simple past sentences ($M = 210.65$, $SD = 41.01$). This tendency to drop a character closer to the path in the past progressive condition, and closer to the destination in the simple past condition, replicates previous evidence that the ongoing nature implied by a past progressive sentence draws attention to the middle portion of the path, whereas there is a tendency to focus attention on the destination in response to simple past sentences.

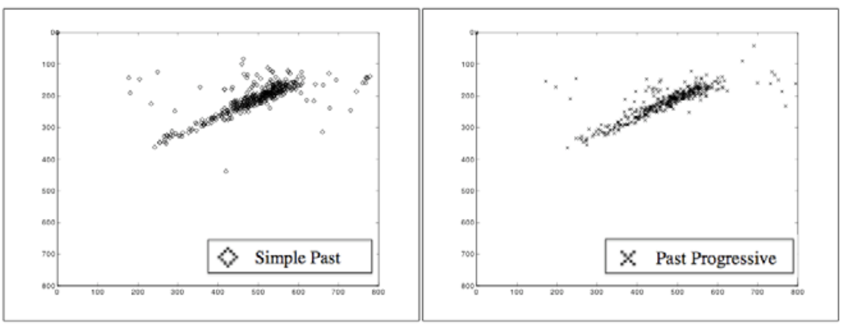


Fig. 2. Drop locations in response to simple past sentences (left panel) and past progressive sentences (right panel)

3.2 Movement Durations

We began our investigation of online processing by looking at the temporal dynamics of the movement of the character. There was no significant interaction of context and aspect when comparing overall movement durations (i.e., the length of time from the initial grab of the character to the final drop of the character into the scene), p's> .2. However, there was a significant interaction of context and aspect on movement durations within in the region of the screen corresponding to the depicted path, $F(1, 63) = 4.6$, $p < .036$. See Figure 3. In the region of the path, the average movement duration for simple past sentences was not substantially different when the context was first described as rough (*M*= 2448.33, *SD*=1848.88) or smooth (*M*=2478.72, *SD*= 1527.17). On the other hand, the average movement duration in the region of the path for the past progressive sentences was slower when the context was first described as rough (*M* = 2667.70, *SD*=1679.86) than when it was described as smooth (*M*=2121.88, *SD*=1240.13). Because simple past sentences focus attention on completed action, context descriptions do not significantly impact the movement dynamics. On the other hand, because past progressive sentences encourage attention to the ongoing-ness of the action, context descriptions of the location of that ongoing action do influence processing. These data extend previous research, suggesting that aspect influences the real-time movement dynamics of the event being described, not merely the endstate of that event. Also, as predicted, the context descriptions modulate this on-line measure when aspect focuses attention to the ongoing action of the motion event.

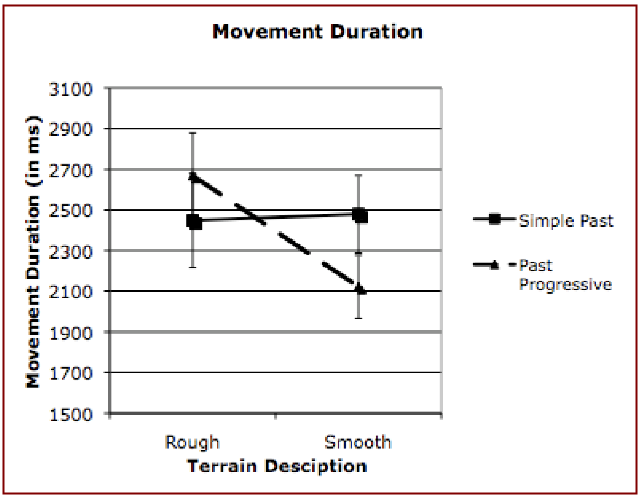


Fig. 3. Movement duration differences in the region of the visual scene corresponding to the path. When participants heard rough terrain descriptions preceding the past progressive sentences, they moved the character more slowly in the region of the screen corresponding to the path than when they heard the terrain described as easy. Terrain description did not change such movement durations when participants heard simple past sentences.

3.3 Raw Time

To begin to compare spatial attraction to the visual scene's path across conditions, we first looked at average x- and y-coordinates within eight 500ms time-bins of the movement duration. There was no significant interaction between aspect and terrain, *p*'s>.1, or main effect of either variable, *p*'s>2. However, breaking the movement into time bins serves only as an approximation of actual attraction over raw time. These 500ms time-bins were not time locked to the sound files, and hence did not have a fixed starting time. Because the offset of the verb occurred late within the sound files and because many participants did not begin to move the character until after the end of the sound file (with an average 1400 ms lag between offset of verb and end of sentence), these data are not synchronized to a fixed point. Future work will address potential raw time spatial differences more fully.

3.4 Spatial Attraction

The movement duration differences only in the region of the path suggest that there is an interaction between the linguistic and visual information. In other words, these movement duration differences are only observed in the relevant region of the visual scene. In order to further explore this apparent interaction of grammatical aspect and visual scene information, we examined the spatial differences in trajectories. Figure 4 shows the average time-normalized trajectories in each of the four conditions. The mean simple past and past progressive trajectories at each of the 101 time-steps in the top panel of Figure 4 illustrate that in the rough terrain context, the average past progressive trajectory curved more toward the path than the average trajectory elicited by the simple past sentences, but only near the end of the trajectory. However, in the smooth terrain description, (Figure 2, bottom), there appears to be greater attraction toward the path across a greater portion of the trajectory for the past progressive sentences.

To determine whether the divergences observed across the simple past and past progressive sentence trajectories in the rough and smooth terrain descriptions were statistically reliable, we conducted a series of t-tests. These analyses were conducted separately on the x- and the y-coordinates at each of the 101 time-steps. In order to avoid the increased probability of a Type-1 error associated with multiple t-tests, and in keeping with Bootstrap simulations of such multiple t-tests on mouse trajectories [14], an observed divergence was not considered significant unless the coordinates between the simple past- and past progressive-sentence trajectories elicited p-values $< .05$ for at least eight consecutive time-steps.

In the rough context description condition, there was significant divergence of the past progressive x-coordinates away from the simple past x-coordinates and toward the path between time-steps 89 and 101, p's $<.04$, and no significant divergence in the y-coordinates. This difference is commensurate with the observed differences in drop locations for past progressive and simple past sentences described earlier. Even though there was no significant interaction between aspect and context description on drop location, this significant divergence so late in the time-normalized trajectories may simply be an artifact of drop locations.

On the other hand, in the smooth context description, there were significant divergences of the past progressive x-coordinates away from the simple past x-coordinates towards the path between time steps 48 and 60, p's $<.4$, and again between time steps 65 and 89, p's $< .03$. There was also significant divergence of the average past progressive y-coordinates away from the average simple past y-coordinates and towards the path between time steps 89 and 101. Again, this y-coordinate divergence late in the trajectory may be an artifact of the drop locations in each condition.

These results are encouraging, but they are not as convincing as the path-movement duration results (Figure 3). It is curious that the spatial attraction differences were detected in the smooth context description but not as robustly in the rough context description. Perhaps the visual stimuli used to depict the path simply did not appear to afford difficult or uneven travel, and the incongruence in the linguistic description and the visual appearance of the path hindered the emergence of full spatial differences in this context description. Future work will investigate this possibility.

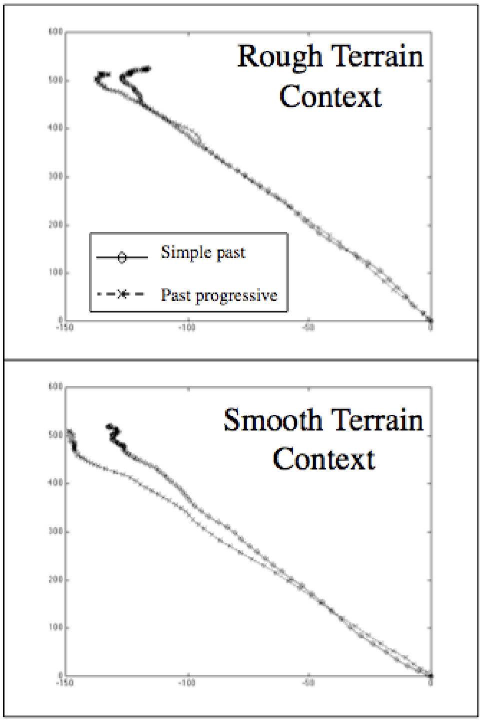


Fig. 4. Average time-normalized simple past and past progressive trajectories in rough and smooth terrain contexts. To some extent in both terrain descriptions (but to a greater extent in the smooth terrain description), participants moved the character into the scene with greater spatial attraction to the path when they had heard a past progressive sentence than when they had heard a simple past sentence.

4 Discussion

Our work provides new insights into how grammar influences the understanding of everyday spatial descriptions, specifically, events that include a path and moving agent. We provide new insights into how different aspectual forms can robustly influence the processing of event descriptions. The results reveal that the processing of spatial descriptions by examining continuous motor output in response to aspectual and contextual differences. Because past progressive aspect focuses attention to the location of the ongoing nature of the verb, contextual descriptions describing the location of the ongoing action significantly interact with grammatical aspect. On the other hand, simple past aspect focuses attention to the location of a completed action, and so contextual description do not influence processing of these sentences in the same way. This is consistent with previous research with mouse-tracking tasks [15], sentence-picture verification tasks [6], and offline spatial reasoning tasks [5].

Although these results are extremely promising, future work is needed. For example, while mouse-tracking is provides a continuous motor output, mouse-tracking is not meant to serve as a general replacement for eye-tracking, specifically because mouse-tracking is not as immediate as the eye-tracking, lagging 200-300ms behind initiating an eye-movement. Therefore, eye-tracking provides more immediate information about intermediate stages of processing. Therefore, future work should look at the immediate time course of processing aspect through the use of more immediate dependent measures. Similarly, our continuing program of research will also look at processing differences with specific verbs, namely verbs that have an inherent perspective to parts of the path, e.g. ‘enter’ and ‘leave’ or ‘come’ and ‘go’.

The current research has implications for several areas of research on spatial language. Although grammatical aspect is known to provide information about the temporal aspects of processing (e.g., completed or not completed, repeated or not repeated), our results indicate that aspect can significantly influence spatial reasoning in the time course of processing. Our work also introduces a novel method for investigating descriptions and depictions of spatial scenes. The beauty of the approach is that it affords careful examination of the temporal dynamics of processing spatial language. In addition, our results provide evidence to support cognitive linguists’ claims regarding meaning as a conceptualization of spatial descriptions, and the idea that aspect, like many domains of language, involves dynamic conceptualization [18, 19].

More broadly, this work resonates with embodied cognition work on perceptual simulation and language understanding [11]. It also dovetails with the methodological advances of Balota and Abrams [20] by providing new evidence from the temporal dynamics of a response after the response has been initiated, and demonstrating that the motor system is not a robot-like automaton triggered by completed cognitive processes. Rather, motor processes are co-extensive with cognitive processes during perceptual/cognitive tasks [e.g., 20, 21]. This work also comports with our understanding of how spatial mental models and visual information are coordinated in motor output. Similarly to the way understanding of spatial events is created and observed through tracking eye movements [17, 22], this work demonstrates that event understanding takes place differently as a function of changes in context descriptions and grammatical aspect. Finally, the work explores a new way that language about space and thought about space are related.

References

1. Spivey, M.J., Grosjean, M., Knoblich, G.: Continuous Attraction Toward Phonological Competitors. PNAS 102, 10393–10398 (2005)
2. Comrie, B.: Aspect. Cambridge University Press, Cambridge (1976)
3. Frawley, W.: Linguistic Semantics. Lawrence Erlbaum Associates, Hillsdale (1992)
4. Croft, W.: Aspectual and Causal Structure in Event Representations. In: Gathercole, V. (ed.) Routes to Language Development: In Honor of Melissa Bowerman, pp. 139–166. Lawrence Erlbaum Associates, Mahwah (2009)
5. Matlock, T., Fausey, C., Cargill, S., Spivey, M.: On the Path Toward Understanding the Dynamics of Aspect Descriptions in Motion Events. Paper presented at 48th Annual Meeting of the Psychonomic Society, Long Beach, California (November 2007)
6. Madden, C.J., Zwann, R.A.: How Does Verb Aspect Constrain Event Representations? Memory & Cognition 31, 663–672 (2003)

7. Ferretti, T.R., Kutas, M., McRae, K.: Verb Aspect and the Activation of Event Knowledge in Semantic Memory. Journal of Experimental Psychology: Learning, Memory, & Cognition 33, 182–196 (2007)
8. Magliano, J.P., Schleich, M.C.: Verb Aspect and Situation Models. Discourse Processes 29, 83–112 (2000)
9. Morrow, D.G.: Prominent Characters and Events Organize Narrative Understanding. Journal of Memory and Language 24, 304–319 (1985)
10. Abrams, R.A., Balota, D.A.: Mental Chronometry: Beyond Reaction Time. Psychological Science 2, 153–157 (1991)
11. Barsalou, L.: Language Comprehension: Archival memory or preparation for situated action? Discourse Processes 28, 61–80 (1999)
12. Barsalou, L.W.: Continuing Themes in the Study of Human Knowledge: Associations, Imagery, Propositions, and Situations. In: Gluck, M.A., Anderson, J.R., Kosslyn, S.M. (eds.) Memory and Mind: A Festschrift for Gordon H. Bower, pp. 209–227. Lawrence Erlbaum Associates, New Jersey (2007)
13. Pulvermuller, F.: Words in the Brain's Language. Behavioral and Brain Sciences 22, 253–279 (1999)
14. Dale, R., Kehoe, C., Spivey, M.: Graded Motor Responses in the Time Course of Categorizing Atypical Exemplars. Memory and Cognition 35, 15–28 (2007)
15. Anderson, S.E., Matlock, T., Fausey, C., Spivey, M.J.: On the Path to Understanding Online Processing of Grammatical Aspect. In: Proceedings of the 30th Annual Conference of the Cognitive Science Society, pp. 143–148. Lawrence Erlbaum Associates, Mahwah (2008)
16. Matlock, T.: Fictive Motion as Cognitive Simulation. Memory & Cognition 32, 1389–1400 (2004)
17. Richardson, D.C., Matlock, T.: The Integration of Figurative Language and Static Depictions: An Eye Movement Study of Fictive Motion. Cognition 102, 129–138 (2007)
18. Langacker, R.W.: Foundations of Cognitive Grammar: Theoretical Prerequisites. Stanford University Press, Stanford (1987)
19. Talmy, L.: Toward a Cognitive Semantics. MIT Press, Cambridge (2000)
20. Balota, A.D., Abrams, R.A.: Mental Chronometry: Beyond Onset Latencies in the Lexical Decision Task. Journal of Experimental Psychology: Learning, Memory, & Cognition 21, 1289–1302 (1995)
21. Gold, J.I., Shadlen, M.N.: Representation of Perceptual Decision in Oculomotor Commands. Nature 404, 390–394 (2000)
22. Spivey, M., Geng, J.: Oculomotor Mechanisms Activated by Imagery and Memory: Eye Movements to Absent Objects. Psychological Research 65, 235–241 (2001)

Implicit Spatial Length Modulates Time Estimates, But Not Vice Versa

Roberto Bottini[1,2] and Daniel Casasanto[2,3]

[1] University of Bergamo, Department of Human Sciences, Bergamo, IT
[2] Max Planck Institute for Psycholinguistics, Neurobiology of Language Group, Nijmegen, NL
[3] Donders Center for Brain, Cognition, and Behavior, Radboud University, Nijmegen, NL
roberto.bottini@unibg.it
daniel.casasanto@mpi.nl

Abstract. How are space and time represented in the human mind? Here we evaluate two theoretical proposals, one suggesting a symmetric relationship between space and time (ATOM theory) and the other an asymmetric relationship (metaphor theory). In Experiment 1, Dutch-speakers saw 7-letter nouns that named concrete objects of various spatial lengths (*tr.* pencil, bench, footpath) and estimated how much time they remained on the screen. In Experiment 2, participants saw nouns naming temporal events of various durations (*tr.* blink, party, season) and estimated the words' spatial length. Nouns that named short objects were judged to remain on the screen for a shorter time, and nouns that named longer objects to remain for a longer time. By contrast, variations in the duration of the event nouns' referents had no effect on judgments of the words' spatial length. This asymmetric pattern of cross-dimensional interference supports metaphor theory and challenges ATOM.

Keywords: ATOM, Metaphor, Psychophysics, Space, Time.

1 Introduction

Space and time are intimately related in the human mind, as they are in the physical world. But exactly *how* are these dimensions related? Here we evaluate two theoretical proposals, one suggesting a symmetric and the other an asymmetric relationship between space and time.

According to the first proposal, space and time are represented in the brain and mind by a common analog magnitude system, which also generates representations of number and quantity. This view, summarized in Walsh's ATOM (A Theory of Magnitude; [1]), is consistent with neurological data showing shared brain areas for processing space, time, and quantity [e.g. 2], and with many behavioral studies in animals and humans [e.g., 3, 4, 5, 6].

Implicit in ATOM is an assumption that these 'ATOMic' dimensions are symmetrically interrelated: not hierarchically related in the brain/mind. Accordingly, Walsh [1] frames predictions in symmetrical terms, positing "overlapping brain regions" and "cross-domain, within-magnitude priming" between dimensions, without specifying

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 152–162, 2010.

any *directionality* to the priming (or interference) effects. Indeed, if space and time are both manifestations of the same general-purpose analog magnitude system, there may be no *a priori* reason to posit that one dimension should depend asymmetrically on another.

On an alternative proposal, space, time, and quantity are importantly related, but in a different way. According to theories of metaphorical mental representation [e.g., 7], representations of time, number, and quantity depend asymmetrically on representations of space. The claim that some domains are asymmetrically dependent on others, which is at the core of metaphor theory, was originally supported by patterns in metaphorical language. In English, it is nearly impossible to talk about domains like time without using words whose primary meaning is spatial (denotatively, developmentally, or historically [8]). Vacations can be *long* or *short*, meetings can be *moved forward* or *pushed back*, deadlines can loom *ahead* or lie *behind* us. Yet, it is far less common to use temporal words to talk about space [7]. This asymmetry in language has been echoed by behavioral findings in psycholinguistics [9], cognitive development, [10], and psychophysics [11].

In one set of studies by Casasanto & Boroditsky, participants viewed lines of various spatial lengths that appeared on a screen for varying durations [11]. They were asked to estimate either the duration or the spatial length of each line, using mouse clicks. Participants were unable to ignore irrelevant spatial information when making judgments about duration, but not the converse. For stimuli of the same average duration, lines that extended shorter in space were judged to take a shorter time, and lines that extended longer in space were judged to take a longer time. By contrast, for stimuli of the same average spatial length, spatial estimation was not affected by the line's duration. This cross-dimensional asymmetry, predicted based on patterns in language, was shown here in non-linguistic psychophysical judgments. Five follow-up experiments varied the attentional, mnemonic, and perceptual demands of the stimuli, and all six experiments supported the same conclusion: mental representations of time depend on representations of space, more than vice versa.

This robust space-time asymmetry supports metaphor theory, but presents a challenge to ATOM. If space and time are both derived from (or are both manifestations of) a general-purpose magnitude metric, then why should representations of time depend on representations of space more than the other way around -- in adults and children, and in language and thought?

It might be possible to reconcile these results with ATOM by positing that in previous studies, space influenced time asymmetrically because space was either (a) the more discriminable dimension, or (b) the more perceptually salient dimension in the stimulus. Discriminability, in this context, refers to the resolution at which a dimension is sampled. Salience means the extent to which one dimension attracts attention relative to the other. Differences in discriminability and perceptual salience have been shown to modulate the strength or direction of cross-dimensional interference and priming effects across numerous studies [12]. In general, the dimension that is more discriminable or salient interferes with the dimension that is less discriminable or less salient. Can task-related differences in the relative discriminability or salience of stimulus dimensions account for the space-time asymmetries observed previously?

One set of studies reviewed above addressed these questions. Tests of cross-dimensional relationships often manipulate more levels of one dimension than of the other, creating an imbalance in discriminability [13]. In the space-time experiments by Casasanto & Boroditsky [11], however, there were 9 levels of each dimension fully crossed, to equate discriminability.

Differences in discriminability may correspond to differences in the accuracy, precision, or variability of judgments across domains. This complicates the interpretation of cross-dimensional interference effects. In the limit, if performance in one domain is perfect, there is no opportunity for variation in the other domain to influence it: the 'clean' domain can influence performance in the 'messy' domain, but not vice versa. In Casasanto & Boroditsky's studies, however, within-domain performance was equivalent across space and time [see also 10].

But is it possible that space was more salient than time in these studies? Following Garner [14], Casasanto & Boroditsky [11] asked participants to judge different dimensions of the same stimuli (e.g., the spatial or temporal extent of a line). Thus, people had the exact same perceptual input during space and time judgments. But this does not guarantee that the dimensions were equally perceptually salient: it is possible to *see* the spatial extent of a line, but not its duration. To address the concern that space may have been more salient than time, in one experiment each line was accompanied by a tone, which sounded for the duration that the line remained on the screen. Tones have temporal extent but no spatial extent. Thus, temporal information was available to the participant through two sensory channels, but spatial information through only one. Yet, increasing the salience of temporal information did not diminish the space-time asymmetry.

Still, on a skeptical interpretation, these previous studies may not have ruled out cross-dimensional differences in perceptual salience definitively. It is possible that space will *always* be more perceptually salient than time whenever perceptible spatial stimuli are used, since it is possible to perceive space, but arguably it is not possible to perceive time directly through the senses [15]. The question remains, then, whether the space-time asymmetry would persist in psychophysical judgments if differences in the perceptual salience of space and time in the stimulus were eliminated.

In the present study, we eliminated differences in perceptual salience by eliminating perceptible variation in the critical dimension (space or time), altogether. We tested whether the implicit spatial information encoded in object nouns can influence estimates of time (in Experiment 1), and whether the temporal information encoded in event nouns can influence estimates of spatial length (in Experiment 2). Participants saw words presented one at a time and reproduced either the duration for which they remained on the screen or their spatial length, using mouse clicks as in Casasanto & Boroditsky [11]. In the duration estimation task (Experiment 1), the target words named objects of various spatial lengths (e.g., *pencil*, *clothesline*, *footpath*). All target words had the same number of letters in Dutch, and therefore the same physical length on the screen. In the spatial length estimation task (Experiment 2) the target words named events of various durations (e.g., *blink*, *party*, *season*). Again, all target words had the same number of letters, but they were presented with a varying number of spaces between letters (1-9 spaces), stretching them out to different spatial lengths on the screen.

Word meanings were irrelevant to the length and duration estimations. We expected, however, that participants would read the words while viewing them, and activate their meanings (voluntarily or involuntarily). Presumably, the meaning of an object noun typically includes a representation of the object's spatial form, and the meaning of an event noun a representation of the event's duration. If internally generated spatial and temporal representations cued by words are sufficient to modulate estimates of experienced duration and spatial length, then we should observe cross-dimensional interference. Following metaphor theory, we predicted that the cross-dimensional interference should be asymmetric, even in the absence of cross-dimensional differences in perceptual salience: spatial representations cued by object nouns should modulate estimates of their duration more than temporal representations cued by event nouns modulate estimates of their spatial extent on the screen.

2 Does Implicit Spatial Length Modulate Time Estimates?

Experiment 1 tested whether the spatial length of a word's referent can modulate estimates of how much time the word remained on the screen.

2.1 Methods

Participants. Native Dutch-speakers (N=39) performed Experiment 1 in exchange for payment.

Materials. Dutch nouns naming 9 concrete objects (Targets) and 9 abstract entities (Fillers) were presented on a computer monitor (resolution = 1024 x 768 pixels) for varying durations. The concrete nouns referred to objects whose characteristic spatial lengths ranged from short (normally measured in centimetres) to long (normally measured in kilometres). English equivalents of these nouns are listed here in order of increasing length: *cigarette*, *pencil*, *ruler*, *meter stick*, *bench, clothesline*, *footpath*, *lane*, *highway.* In Dutch, all 9 target nouns had 7 letters, and were presented on the screen in a fixed-width font (62-point Courier New). Therefore, the targets did not differ in their physical spatial lengths on the screen; rather, they differed in their implicit lengths (i.e., the typical spatial lengths of their referents).

The filler nouns referred to abstract entities that have no physical spatial length: *guess*, *idea*, *pride*, *opinion*, *envy*, *thought*, *philosophy*, *suspicion, dignity*. However, they varied in their number of letters in Dutch (from 3-11 letters) and therefore in their physical length on the screen (nine different lengths, varying from 50-450 pixels as measured from the left edge of the first letter to the right edge of the last letter). By contrast with the targets, the fillers did not differ in the implicit lengths of their referents; rather, they differed in their physical lengths on the screen.

Each target and filler word was presented 9 times throughout the experiment, for 9 different durations. Durations ranged from 1000 to 5000 ms in 500 ms increments. Fully crossing these 9 durations with the target words (which had 9 different implicit spatial lengths) produced 81 target trials. Likewise, fully crossing the 9 durations with the filler words (which had 9 different physical lengths on the screen) produced 81 filler trials. The 162 different trials were presented in random order, with fillers and targets intermixed. Words were presented in white letters on a black background in the center of the screen. Participants were tested individually and testing lasted about 30 minutes.

Procedure. Participants viewed the 162 words, one word at a time, from a viewing distance of approximately 50 cm. Immediately after each word disappeared an "hourglass" icon appeared in the upper left corner of the monitor indicating that the subject should reproduce the amount of time the word remained on the screen. To estimate duration, subjects clicked the mouse once on the center of the hourglass, waited the appropriate amount of time, and clicked again in the same spot, thus indicating the beginning and end of the temporal interval. All responses were self-paced. After the experiment there was a two-part debriefing. In the first part, the experimenter asked the participant "What do you think this experiment is about?" and "What do you think we were looking for?" to determine whether the participant was aware of any relationship between the implicit lengths of the target words and their durations. In the second part, participants saw each target word again, in random order, and verbally estimated the typical spatial length of the target words' referents (using an appropriate unit of measurement). These subjective length estimates were used in later analyses as predictors of subjective duration.

2.2 Results and Discussion

Four participants were removed from the analyses below: one for giving nonsensical answers in the debriefing, one for excessively poor time estimation performance according to the criterion used by Casasanto & Boroditsky [11][1], and two for guessing that there was a connection between the meanings of the target words and time estimation.

For the remaining 35 participants, we first analyzed participants' duration estimates as a function of the actual duration of the stimuli. Overall, duration estimates for target words were highly accurate (mean effect of actual duration on estimated duration: $y=0.83x + 154.11$, $r^2=.99$, df=7, $p<.001$; fig 1a).

We then tested for effects of implicit length on duration estimation. Target words were rank-ordered according to the typical lengths of their referents (this *a priori* ranking was confirmed by participants' post-test length estimates). Non-parametric correlation showed that implicit spatial length affected estimates of duration ($y=3.77x + 2605.70$, $r_{s(\text{Spearman's rho})}=0.75$, df=7, $p<.001$; fig.1b).

Finally, we conducted a parametric analysis of the effect of implicit length on duration estimation. Participants' post-test ratings of the typical spatial length of each target word's referent were used as a predictor of their duration estimates. Ratings for each target item were averaged, and the average length estimates in meters were transformed by a base 10 logarithm. This analysis corroborated the non-parametric analysis, showing a highly significant effect of implicit spatial length on duration estimation ($y=5.60x + 2619.20$, $r^2=.57$, df = 7, $p< .001$).

Participants incorporated irrelevant spatial information into their temporal estimates. For stimuli of the same average duration, words with (spatially) shorter referents were judged to remain on the screen for a shorter time, and words with longer referents for a longer time. This was true even though the task did not require participants to process the words' meanings.

---

[1] Participants were excluded if the slope of their within-domain duration or length estimates was less than 0.5 [see, 11]. This criterion, which resulted in the exclusion of only one participant overall, is unbiased with respect to the predicted cross-dimensional interference because length and duration are orthogonal in the designs of both experiments.

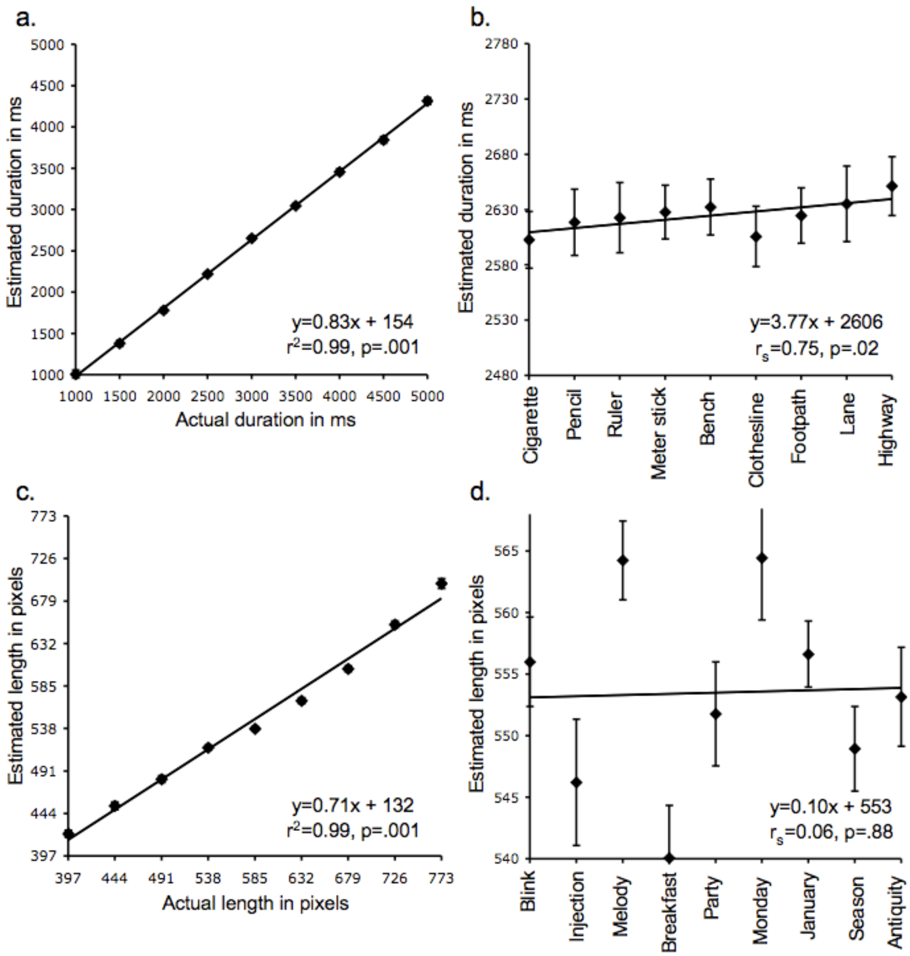


Fig. 1. Results of Experiment 1 (top) and Experiment 2 (bottom). 1a. Within-domain effect of actual word duration on estimated duration. 1b. Cross-domain effect of words' implicit spatial length on estimated duration. 1c. Within-domain effect of actual word length on estimated spatial length. 1d. Cross-domain effect of words' implicit duration on estimated spatial length. The axes of the top and bottom plots (a-c, b-d) are proportional with respect to the total range of target values. Error bars show s.e.m.

This result shows that perceptible spatial input is not necessary to modulate time estimates; rather, internally- generated spatial representations cued by words are sufficient. This outcome, *per se*, is equally consistent with metaphor theory and with ATOM. To distinguish between the theories, it is necessary to conduct a complementary experiment to determine whether implicit duration can affect estimates of spatial length, and whether cross-dimensional interference effects are as symmetric, as expected on ATOM (Effect of Space on Time ≈ Effect of Time on Space) or asymmetric, as predicted by metaphor theory (Effect of Space on Time > Effect of Time on Space).

3 Does Implicit Duration Modulate Estimates of Spatial Length?

Experiment 2 tested whether the duration of a word's referent can modulate estimates of the word's spatial length as presented on the screen.

3.1 Methods

Participants. Native Dutch-speakers (N=35) performed Experiment 2 in exchange for payment.

Materials. Dutch nouns naming 9 events (Targets) and nine concrete objects (Fillers) were presented on a computer monitor (resolution = 1024 x 768 pixels). The target nouns referred to events whose characteristic durations ranged from short (normally measured in seconds) to long (normally measured in years). English equivalents of these nouns are listed here in order of increasing duration: *blink*, *injection*, *melody*, *breakfast*, *party*, *Monday*, *January*, *Season*, *Antiquity.* All targets were presented for 3000ms. Therefore, the targets did not differ in the physical durations for which they remained on the screen; rather, they differed in their implicit durations (i.e., the typical durations of their referents).

The filler nouns referred to concrete objects that have no inherent duration: *doormat*, *ballast*, *portrait*, *detritus*, *crystal*, *device*, *case*, *sawdust, handle*. Each filler noun appeared for 9 different durations from 1000-5000ms, increasing in 500ms increments. By contrast with the targets, the fillers did not differ in the implicit durations of their referents; rather, they differed in the physical durations for which they remained on the screen.

In Dutch, all target and filler nouns had seven letters, and were presented on the screen in a fixed-width font (62-point Courier New). Each word was presented 9 times throughout the experiment, with a varying number of spaces in between the letters (1-9), to stretch the words out to 9 different spatial lengths on the screen. Due to the font selected, word lengths ranged from 397 to 773 pixels, in 47 pixels increments. Presenting each word at each of these 9 spatial lengths produced 81 filler trials and 81 target trials. For the fillers, spatial length was fully crossed with the physical duration for which they were presented. For the targets, spatial length was fully crossed with the implicit duration of their referents. The 162 different trials were presented in random order, with fillers and targets intermixed. Words were presented in white letters on a black background in the center of the screen. Participants were tested individually and testing lasted about 30 minutes.

Procedure. Participants viewed the 162 words, one word at a time, from a viewing distance of approximately 50 cm. Immediately after each word disappeared an "X" appeared in the upper left corner of the monitor indicating that the subject should reproduce the spatial length that the word had occupied on the screen. To estimate length, subjects clicked the mouse once on the center of the X, moved the mouse to the right the appropriate distance, and clicked again, thus indicating the beginning and end of a spatial interval. All responses were self-paced. After the experiment there was a two-part debriefing, as in Experiment 1. The first part was to determine whether the participant was aware of any relationship between the implicit durations of the target words and their spatial lengths. In the second part, participants saw each target word again, in random order, and verbally estimated the typical duration of the target words' referents (using an appropriate unit of measurement). These subjective duration estimates were used in later analyses as predictors of subjective spatial length.

3.2 Results and Discussion

One participant was removed from the analyses below for guessing that there was a connection between the meanings of the target words and spatial length estimation.

For the remaining 34 participants, we first analyzed participants' spatial length estimates as a function of the actual spatial length of the stimuli. Overall, length estimates for target words were highly accurate (mean effect of actual length on estimated length: $y=0.71x + 132.44$, $r^2=.99$, df=7, $p=.001$; fig 1c).

We then tested for effects of implicit duration on spatial length estimation. Target words were rank-ordered according to the typical durations of their referents (this *a priori* ranking was confirmed by participants' post-test duration estimates). Non-parametric correlation showed that implicit duration did not affect estimates of spatial length ($y=0.10x + 553.00$, $r_{s(\text{Spearman's rho})}=0.06$, df = 7, *ns*; fig.1d).

Next, we conducted a parametric analysis using participants' post-test ratings of the typical duration of each target word's referent were used as a predictor of their length estimates. Ratings for each target item were averaged, and the average duration estimates in minutes were transformed by a base 10 logarithm. Again, there was no effect of implicit duration on spatial length estimation ($y= 0.04x + 553.39$, $r^2=.0003$, df=7, *ns*).

Finally, we compared the strength of the cross-dimensional interference effects across Experiments 1 and 2. The difference of correlations showed the predicted cross-dimensional asymmetry ($r_{\text{effect of spatial length on duration}}-r_{\text{effect of duration on spatial length}}=0.74$, $z=1.66$, $p=0.05$, one-tailed; see fig. 1b, 1d). This difference cannot be attributed to differences in within-domain performance ($r_{\text{effect of actual duration on estimated duration}}-r_{\text{effect of actual spatial length on estimated spatial length}}=0.00$, $z=0.00$, *ns*; see fig. 1a, 1c).

4 General Discussion

This study tested whether implicit spatial information encoded in concrete object nouns can influence estimates of time (in Experiment 1), and whether implicit temporal information encoded in event nouns can influence estimates of spatial length (in Experiment 2). When participants reproduced the duration for which an object noun remained on the screen, their estimates were influenced by the implicit length of the word's referent. Words that named shorter objects (e.g., *cigarette*, *pencil*) were judged to last a shorter time, and words that named longer objects (e.g., *bench*, *highway*) to last a longer time. By contrast, when participants reproduced the spatial length of an event noun, the duration of the word's referent did not influence judgments of spatial length.

This asymmetric pattern of cross-dimensional interference was predicted based on patterns in language: people talk about time in terms of space more than they talk about space in terms of time [7]. These data show that people incorporate spatial information into their temporal judgments even when they're not using any metaphorical language, and support the hypothesis that mental representations of time are asymmetrically dependent on representations of space: people use spatial length to think about duration, more than vice versa.

This space-time asymmetry cannot be attributed to differences in how well participants reproduced the actual durations and lengths of the stimuli, *per se*, since there was no significant difference between the effect of actual duration on estimated duration (fig. 1a) and the effect of actual length on estimated length (fig. 1c). Thus, differences in cross-dimensional interference were not due to differences in within-domain performance.

Furthermore, the space-time asymmetry cannot be attributed to differences in the perceptual salience of the interfering dimensions (i.e., space in Expt. 1, time in Expt. 2). In previous experiments, space could have influenced time asymmetrically because space is inherently more perceptually salient than time (which some scholars have argued can never be perceived directly [15]). But here there was no perceptible variation in the spatial component of duration-reproduction stimuli, and no perceptible variation in the temporal component of length-reproduction stimuli. Internally generated representations of spatial length, cued by words, were sufficient to modulate estimates of the words' physical duration. This was true even though the words' meanings were task-irrelevant.

Before discussing theoretical implications of these data further, it is important to consider whether the observed pattern could be due to unintended features of the stimulus words. For example, is it possible that duration estimates in Experiment 1 were influenced by implicit *speed* encoded in the concrete nouns, rather than implicit length? The three longest objects (*footpath*, *lane*, and *highway*) are all spatial paths. The speed of motion associated with these paths increases with their lengths (i.e., *footpath-walking*, *lane-slow driving*, *highway-fast driving*). The conflation of length and speed in these items was a consequence of restrictions on the stimuli: items had to increase in ordinal length unambiguously, and had to have 7 letters in Dutch.

If the effect of object length on duration estimates had been driven by these three items, this would be problematic. However, even a causal inspection of fig. 1b shows this was not the case. For the majority of the items there were no clear speed associations, and yet the effect of implicit length was found. For the first 5 items (*cigarette*, *pencil*, *ruler*, *meter stick*, *bench*), ordinal increases in implicit length corresponded to a monotonic increase in estimated duration. The predicted effect of length on duration was significant in these 5 items, alone ($y=6.84x + 2600$, $r_{s(\text{Spearman's rho})}=1.00$, $p=.001$). Thus, implicit speed was not responsible for the effect of implicit spatial length we report here (see [11], Expt. 6 for further evidence that spatial length affects duration estimates independent of speed).

On another skeptical possibility, could implicit *duration* encoded in object nouns have produced the observed effect on duration estimation? Looking at the longest and shortest items alone, this seems plausible. *Cigarette* could be associated with the time it takes to smoke a cigarette (a short time), and *highway* with the amount of time one typically drives on a highway (a longer time). Yet, looking at the full range of stimuli, this alternative explanation seems implausible. What durations are prepotently associated with *clothesline*, *pencil*, *ruler*, *bench*, or *meter stick*? Ordinal increases in spatial length predicted ordinal increases in duration estimates for 7 out of the 8 ordinal pairs of stimuli (i.e., *cigarette* < *pencil*; *pencil* < *ruler*; *ruler* < *meter stick*; etc.) Pairwise differences in the typical spatial lengths of the words' referents are self-evident (and were confirmed by participants' post-test ratings), but for most of these word pairs, it seems unlikely that there are corresponding pairwise differences in durations associated with the words' referents.

Finally, although the space-time asymmetry cannot be due to differences in the *perceptual salience* of the interfering dimensions, could they be due to differences in *conceptual salience*? Could the spatial component of the object words' meanings be more salient than the temporal component of the event words' meanings? We cannot rule out this possibility definitively, but this seems unlikely to be the case. It is difficult to evaluate how salient spatial length is in the meaning of *bench* or *cigarette*, and to compare this with the salience of temporal duration in the meaning of *melody* or *party.* But a few of the stimuli are very strongly associated with a unit of space (*ruler*, *meter stick*) or a period of time (*Monday*, *January*, *season*, *Antiquity*). For these items, it is reasonable to assume that a spatial or temporal representation is the most salient aspect of the word's meaning. This was the case for only two of the object words (22% of targets) but for four of the event words (44% of targets). Therefore, overall, it seems likely that any asymmetry in conceptual salience favored the temporal meanings of the event words, thus working against the hypothesized space-time asymmetry.

These results suggest that the asymmetric dependence of time on space in psychophysical judgments is not an artifact of perceptual or conceptual asymmetries built into the stimuli. Rather, this performance asymmetry reflects a fundamental difference in the way people mentally represent space and time. Yet, this asymmetric relationship between space and time in the mind may, indeed, result from an asymmetry in how perceptible space and time are more broadly -- not in any particular experimental stimuli, but rather in the observable world, in general. Space and time are correlated in our everyday experiences (e.g., as objects travel farther more time passes), and tracking these correlations may be useful for anticipating changes in the physical environment. Correlation is a symmetrical relationship, but people may rely more heavily on the more perceptually available dimension (space), using it heuristically as an index of changes in the less perceptible dimension (time).

It appears that time and space are, in Garner's [15] terminology, *asymmetrically separable* dimensions: it is possible to ignore irrelevant variation in time while judging space but not possible (or more difficult) to ignore irrelevant variation in space when judging time. At present, there is nothing in Walsh's [1] ATOM proposal that can predict or explain the asymmetric separability of space and time. Yet, this cross-dimensional relationship is readily predicted by metaphor theory.

Importantly, space and time are predicted to be related *asymmetrically* but not *unidirectionally.* There is evidence that time can influence space in some paradigms [e.g., 11], just as people can sometimes use temporal words to talk about space (e.g., "*I live two minutes from the station*" is a temporal metaphor for spatial distance). Simply showing that time can influence spatial judgments in some cases does not challenge the asymmetry we report here: to address the question of asymmetry, the cross-dimensional influences of time and space must be appropriately compared, controlling for salience and discriminability across dimensions, and for within-dimension performance.

We propose that Garner-like tests of dimensional separability will be critical for either modifying ATOM or deciding to abandon it in favor of a metaphorical theory of spatial, temporal, and numerical magnitude representation. In order to understand how space, time, and other prothetic dimensions are represented in the brain and mind, it is necessary to go beyond investigating *whether* these dimensions interact and determine *how* they interact.

References

1. Walsh, V.: A theory of magnitude: common cortical metrics of time, space and quantity. Trends in Cog. Sci. 7(11), 483–488 (2003)
2. Basso, G., Nichelli, P., Frassinetti, F., di Pellegrino, G.: Time perception in a neglected space. Neuroreport 7, 2111–2114 (1996)
3. Church, R.M., Meck, W.H.: The numerical attribute of stimuli. In: Roitblat, H.L., Beaver, T.G., Terrace, H.S. (eds.) Animal Cognition, pp. 445–464. Erlbaum, Mahwah (1984)
4. Fischer, M.: Cognitive representation of negative numbers. Psych. Sci. 14(3), 278–282 (2003)
5. Gallistell, R.C., Gellman, R.: Non-verbal numerical cognition: from reals to integers. Trends in Cog. Sci. 4, 59–65 (2000)
6. Cappelletti, M., Freeman, E.D., Cipolotti, L.: Dissociations and interactions between time, numerosity and space processing. Neuropsychologia 47(13), 2732–2748 (2009)
7. Lakoff, G., Johnson, M.: Philosophy in the flesh: The embodied mind and its challenge to Western thought. University of Chicago Press, Chicago (1999)
8. Clark, H.H.: Space, time, semantics and the child. In: Moore, T.E. (ed.) Cognitive Development and the Acquisition of Language, pp. 27–63. Academic Press, New York (1973)
9. Boroditsky, L.: Metaphoric structuring: understanding time through spatial metaphors. Cognition 75(1), 1–28 (2000)
10. Casasanto, D., Fotakopoulou, O., Boroditsky, L.: Space and Time in the Child's Mind: Evidence for a Cross-Dimensional Asymmetry. Cog. Sci. 34, 387–405 (2010)
11. Casasanto, D., Boroditsky, L.: Time in the Mind: Using space to think about time. Cognition 106, 579–593 (2008)
12. Santiago, J., Román, A., Ouellet, M.: Flexible foundations of abstract thought: A review and a theory (submitted)
13. Pansky, A., Algom, D.: Stroop and Garner effects in comparative judgement of numerals: The role of attention. JEP:HPP 25, 39–58 (1999)
14. Garner, W.R.: Interaction of Stimulus Dimensions in Concept and Choice Processes. Cog. Psych. 8, 98–123 (1976)
15. Ornstein, R.: On the experience of time. Penguin, Hammondsworth (1969)
16. Miles, L.K., Nind, L.K., Macrae, C.N.: Moving Through Time. Psych. Sci. (2010), doi:10.1177/0956797609359333

Bio-inspired Architecture for Active Sensorimotor Localization

Thomas Reineking*, Johannes Wolter,
Konrad Gadzicki, and Christoph Zetzsche

Cognitive Neuroinformatics, University of Bremen,
P.O. Box 330 440, 28334 Bremen, Germany
trking@informatik.uni-bremen.de

Abstract. Determining one's position within the environment is a basic feature of spatial behavior and spatial cognition. This task is of inherently sensorimotor nature in that it results from a combination of sensory features and motor actions, where the latter comprise exploratory movements to different positions in the environment. Biological agents achieve this in a robust and effortless fashion, which prompted us to investigate a bio-inspired architecture to study the localization process of an artificial agent which operates in virtual spatial environments. The spatial representation in this architecture is based on sensorimotor features that comprise sensory sensory features as well as motor actions. It is hierarchically organized and its structure can be learned in an unsupervised fashion by an appropriate clustering rule. In addition, the architecture has a temporal belief update mechanism which explicitly utilizes the statistical correlations of actions and locations. The architecture is hybrid in integrating bottom-up processing of sensorimotor features with top-down reasoning which is able to select optimal motor actions based on the principle of maximum information gain. The architecture operates on two sensorimotor levels, a macro-level, which controls the movements of the agent in space, and on a micro-level, which controls its eye movements. As a result, the virtual mobile agent is able to localize itself within an environment using a minimum number of exploratory actions.

1 Introduction

In spite of substantial advances in the design of artificial intelligent systems, biological systems still represent the desirable ideal in many contexts. This is also true regarding a basic competence in spatial cognition: the ability to determine one's own location within the environment. In this paper our aim is to investigate how we can make use of results and concepts from psychology and neurobiology in the design of a bio-inspired architecture for vision-based localization.

For this we have to consider several factors: First, a basic prerequisite is an adequate representation of the environment. The notion of a *cognitive map* is often seen as an abstract copy of the physical layout, which may, for example, resemble

* Corresponding author.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 163–178, 2010.

an annotated cartographic representation, which can be substantially distorted with respect to the true metrical and geometrical properties, but which is basically similar to a two-dimensional image-like entity. This concept of a "map in the head" has been repeatedly criticized as potentially misleading, e.g., [17,42]. In our opinion, the most critical shortcoming of this concept is the absent or indirect role of motor actions in the representational model. In most cases, there is a clear separation between configurations of spatial entities and the actions that can be performed on/with it. This is in stark contrast to recent developments in perceptual psychology and neurobiology. Starting with the affordance concept of Gibson [10] over the common coding theory of Prinz [26,14] to the concept of sensorimotor contingencies of O'Regan [23], many psychological theories argue that the strict separation of sensory and motor components in the representational concepts is no longer tenable. The well-studied primate vision also brought up evidence for a stronger coupling of sensory and motor processes. For example, the system of canonical and mirror neurons revealed an intricate coupling of perception and motor control [30], and the postulated dorsal-ventral stream also shows representations of space used for the organization of actions [31]. Likewise, the firing of visual neurons in the ventral path, which were historically ascribed to early and solely sensory processing, turned out to be directly related to eye movements [20].

In understanding how a representation is organized it is also important to consider how it is established under natural behavioral conditions. Typically, this is a dynamic process in which motor actions play an essential role. Mobile agents move within the environment and produce a sequence of motor actions, and each action changes the relation between the agent and the environment. From the static perspective on a spatial representation, this is a disaster, but research in active perception has revealed that these motor actions actually simplify the development of a reliable representation of the environment [1,2]. It should be noted that motor actions also play an important role in the developmental landmark-route-survey (LRS) concept of spatial representation proposed in [37], although the final stage, the survey representation, again represents an image-like cognitive map. Finally, our own experiments with physically "impossible" virtual environments provide strong evidence against the concept of an image-like cognitive map [50,48], which is in line with a number of other studies which also found evidence against a cognitive map in the sense of an enduring allocentric representation [11,43,7]. Taken together, this indicates that a biologically plausible spatial representation should also comprise motor information, and preferably not as a simple add-on but in an integrated combination with sensory information. This leads us to make use of a sensorimotor representation of the spatial environment in our architecture.

A second important point in the design of a biologically plausible architecture is efficiency. Biological system often achieve their goals with a minimization of both effort and use of resources. Regarding information processing, this can be formalized as information-theoretic optimization. For example, the neural processing in the visual system can be successfully described as a result of such an information theoretic optimization (e.g., [49]). For a biologically plausible architecture it would thus be desirable to obtain a maximum amount of information about an environment with a minimum number of motor actions [34]. As a last point in our design considerations, we have to take into account that biological representations are typically not established and used in a purely bottom-up fashion, but are part of an action-perception cycle, which involves bottom-up processing as well as top-down control.

Here, we approach the aforementioned ideas by the design of an artificial system, the Sensori-Motor Explorer (SMX), which is a virtual mobile agent that uses sensorimotor features as basic representational elements for exploratory localization in virtual spatial environments. The system presented here results from an integration of our sensorimotor representation [34,51] with a temporal belief update mechanism [27] and a learning component which allows for the unsupervised learning of the underlying hierarchical sensorimotor representations [9]. Central to the system is the use of the principle of maximum information gain to compute and execute the most informative actions. As a result, the SMX can localize itself in its environment using a minimum of exploratory steps.

The paper is organized as follows. In section 2, we provide a brief overview of the system properties of SMX, of its micro-level and macro-level exploration behavior, and of the generic hybrid architecture that is used to control both levels. The individual components of the system are then explained in more detail in section 3. This section contains also descriptions of the learning mechanism for the generation of sensorimotor hierarchies and of the temporal update of the belief distribution in response to spatial context changes. The resulting system behavior is described section 4. The paper concludes with a discussion of the major achievements.

2 System Architecture

The SMX performs exploration and localization in a VR environment. We use virtual reality and simulation in our research because this provides us with simple and complete control over all properties of both the environment and the agent. In particular, we can easily investigate the influence of features, objects and spatial arrangements on the performance of the system. In the current state, we use indoor environments consisting of rooms which are populated by typical objects like chairs, bookshelves, etc. The objects and the room walls have uniform or simple textures, typically with static lighting conditions.

The agent is characterized by two major features: first, it operates on two behavioral levels with different sensorimotor granularity, and, second, both behavioral levels are controlled by a single hybrid architecture, which integrates bottom-up sensory processing with a top-down uncertainty minimization strategy. The two sensorimotor levels are illustrated in Fig. 1: at the micro-level, a local view of the environment is explored in a detailed analysis by saccadic eye movements. At the macro-level, the agent performs exploratory movements within the spatial environment. In the bottom-up stage, features are extracted from the environment and combined with motor data. The resulting sensorimotor features are the basic representational elements of the system. At the micro-level, these features comprise local image features and the motor data for saccadic eye movements while, at the macro-level, the sensorimotor features combine information about local views with the motor data for changing the agent's spatial location in the environment. At both levels, an estimation of the current state is performed based on the observed sensorimotor features. States at the macro-level are distinctive regions in space (in this particular case rooms) while states at the micro-level correspond to single views, which, in turn, form features at the macro-level. The resulting hypothesis spaces have a tree-like structure where nodes higher-up in the hierarchy represent disjunctions of states. They are obtained by hierarchically clustering a large number of sampled sensorimotor features [9]. Such a hierarchical structure has in fact been identified as a key property of spatial representations [13,45], and the formalization of the resulting estimation problem naturally lends itself to a belief function representation in Dempster-Shafer framework. In particular, this allows the agent to remain agnostic with respect to the belief distribution over the leaf nodes if the sensorimotor evidence does not support specific leaf nodes.

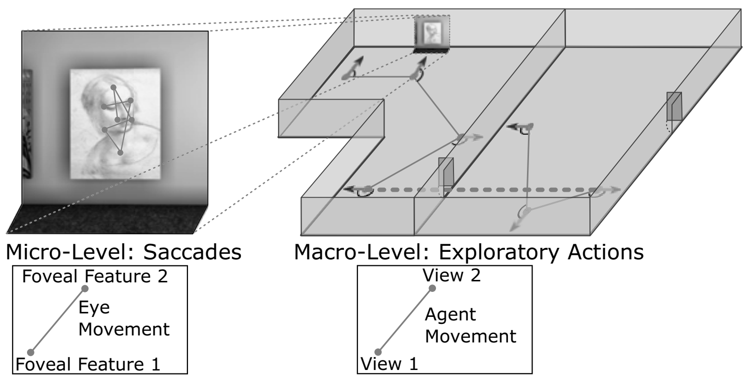


Fig. 1. Two levels of sensorimotor granularity. The micro-level scene analysis (left) is based on saccadic eye movements on a single view. At the macro-level (right), the agent moves between different locations in the environment.

The architecture used at both behavioral levels is shown in Fig. 2. In each action-perception cycle, a new sensorimotor feature is extracted and the belief distribution over the hierarchy of the corresponding level is updated based on a statistical model of state-feature co-occurrences which are learned in an initial training phase. The top-down component uses the updated belief in order to compute the expected information gain associated with subsequent features and their corresponding actions [34]. The sensorimotor levels operate in an interleaved fashion, with results from the micro-level passed as input to the macro-level. Together, they enable the agent to perform a statistically optimal sequence

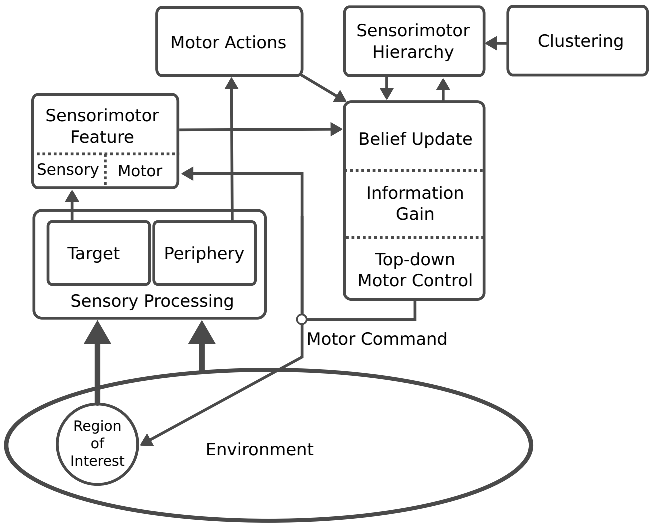


Fig. 2. Generic hybrid architecture. The same type of architecture is used for the micro- and the macro-level control. Sensory input and motor data are combined in the bottom-up processing to obtain a sensorimotor feature, which is used to update the belief distribution over the hierarchy at the corresponding level. From this, the top-down strategy selects a new action by minimizing the expected uncertainty.

of exploratory actions to gain information about the environment. The different components of the system, the internal sequence of operations and the resulting behavior are described in more detail in the following section.

3 Components of the System

3.1 Sensorimotor Features

A characteristic property of the system is its representation of the spatial environment via sensorimotor features, which form the basic representational elements for both the macro-level and the micro-level. A sensorimotor feature is a triple of the form $f = [v_1, a, v_2]$ where v_1 is the sensory feature vector obtained prior to action a and v_2 is the sensory feature vector obtained after executing a. By making actions an explicit part of the representation, they become an additional source of information for the state estimation because one can make use of their correlation with states—a feature not present in classical localization approaches where one is typically agnostic with respect to the question of where certain actions are likely to occur, e.g., [40]. The continuous-valued sensorimotor vector f is mapped to the closest element f_i of a finite set of prototype vectors. At the micro-level, the motor component of each feature is a saccadic eye movement, and the sensory components are derived by a biologically motivated vision system by use of a local wavelet analysis that is applied to the pre- and post-saccadic fixation points.

At the macro-level, the motor component is a movement of the complete agent in space, and the sensory components are labels (for the local views) that the agent registers before and after the movement. The label for a local view is the result of the micro-level analysis of the local view by saccadic eye movements, and this label is then passed as sensory input to the macro-level analysis. Each level makes use of a discrete set of sensorimotor micro- and macro-level features, and they are acquired in an initial exploration process in which the association of sensorimotor features with states in the environment is established by supervised learning.

An action at the macro-level consists of two rotations and one translation. The first rotation turns the agent at the starting location in the direction of the target location, and the following straight movement gets the agent to this location. Here a second rotation aligns the agent to the orientation of the target view. At the macro-level there is a distinction between the handling of intra- and inter-room sensorimotor features. An intra-room feature belongs to a single room and the updating for this case is described in 3.4. Inter-room features, on the other hand, are those where the pre-action part belongs to one room and the post-action part belongs to another. Here, the resulting belief has to be transferred to the corresponding destination, which is described in 3.5.

3.2 Saccadic Eye Movements

An essential problem of processing a visual scene is the detection of the most informative visual regions (e.g., those parts of an object, which are most informative for its identification). Information about the image structure at these few locations is usually sufficient to draw reliable conclusions about the local scene. Biological vision systems have developed an efficient design in which the pattern recognition capabilities are concentrated in a small region of the visual field, the central fovea, whereas the periphery has only limited optical resolution and processing power. With a static eye, one can hence only see a small spot of the environment with satisfactory quality, but this spot can be rapidly moved with fast saccadic eye movements of up to $700°/s$ towards all the "relevant" regions of a scene. This selection process is determined by bottom-up processes on the input scene as well as by top-down processes determined by the memory, internal states and current tasks [47].

In order to enable an efficient selective "sampling" of a local scene by saccadic eye movements, we have integrated bottom-up and top-down processing into a hybrid architecture (cf. Fig. 2). With respect to saccadic scene analysis, the sensory processing stage in this architecture has two functions. On the one hand, the pre-processing stage has to identify highly informative candidate locations within the scene, which can be the target of saccadic fixations, and, on the other hand, it has to provide detailed information about the fixated local pattern. It consists of a wavelet-like image decomposition by size- and orientation-specific filters and by nonlinear saliency operators based on the concept of intrinsic dimensionality. A detailed description of the filters and the non-linear operators can be found in [49] and [34]. In addition, we process each scene by a ratio of Gaussians first in order to increase luminance invariance. The extracted visual features are combined with the motor information necessary for shifting the focus to the next fixation point, thus forming a micro-level sensorimotor feature which is then used to update the belief about the current scene.

3.3 Generating Sensorimotor Hierarchies

The hypotheses used by the SMX are represented in a hierarchical manner for efficiency as well as for coping with non-specific evidence. The macro- and the micro-level both have their individual hierarchical representation. Each node $H \in \mathbf{H}$ in the respective hierarchy $\mathbf{H}$ represents a set of singleton hypotheses, and the leaf nodes represent hypotheses about individual items. The leaf hypotheses are currently pre-defined while the higher-level nodes representing sets of views or rooms are generated in an unsupervised clustering process. Fig. 3 illustrates this structure for both, the micro- and macro-level.

The hierarchical structure in our system is organized according to the similarity of the sensorimotor information associated with different states, i.e., views and rooms sharing similar features are grouped together in the clustering process. In the past, we have investigated alternative grouping principles, in particular

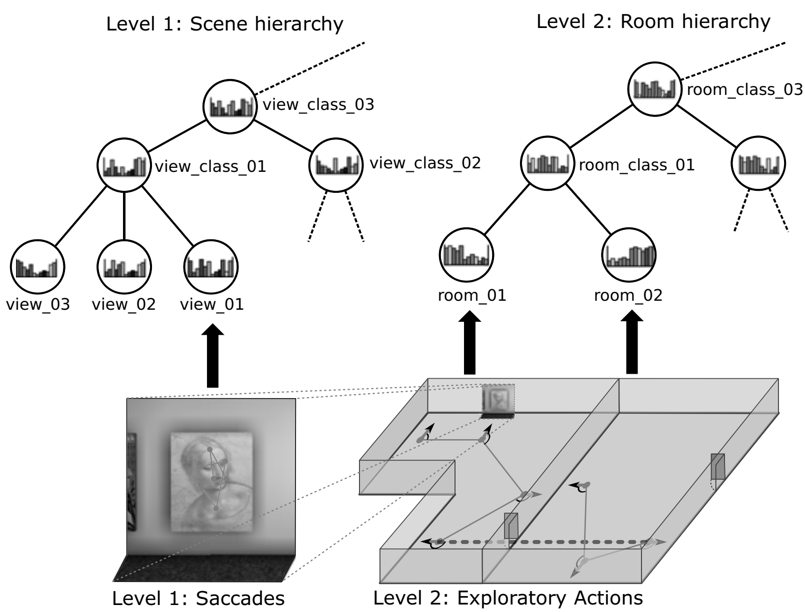


Fig. 3. Two levels of hierarchical sensorimotor representations. At the micro-level (left) scenes are organized in a hierarchical manner with individual scenes as leafs. A histogram of quantized saccadic information is stored in each node as a description of the scene/class of scenes. At the macro-level (right) an equivalent representation is maintained for spatial information with leaf nodes representing single rooms. A room description consists of a histogram of quantized macro-level sensorimotor features (views with motor action).

"spatial similarity" [28], where each node represents a connected region in space, and semantic similarity [29] based on an ontological model. The most suited principle depends on the task, and we use sensorimotor similarity here, because the resulting clusters are the most relevant for scene analysis/localization[1] while a region-based organization might be more appropriate for a task like navigation.

In a previous version of the system [51], these hierarchies were constructed manually for the domain. In order to automate this process, we developed an unsupervised learning method that generates a hierarchy by performing an agglomerative clustering on distributions of sensorimotor features associated with a each hypothesis H. For this, one first needs to measure the similarity of two sensorimotor features. Treating them as symbols and testing for equality worked well for the recognition task described in [34], but it restricts the measure to set-based operations. Initial tests with such similarity measures produced only mediocre results and led us to consider a numerical representation. Similar to approaches from document clustering [32], we thus decided to use a numerical representation which allows for the usage of more suited similarity measures.

For obtaining the numerical representation, sensorimotor features are extracted from a large set of samples and a Self-Organizing Map (SOM) [16] is trained with the entire sample set. That way the system learns how to group sets of features for a given output size (the map size) which corresponds to the number of components of the vector used for the representation of an instance. The actual representation for a particular instance is obtained by processing all sensorimotor features associated with a instance by the SOM. For each feature, a particular map node will be maximally activated. By counting the activations of map nodes for all features, a histogram is generated for each instance, which can then be compared by the similarity measure. Based on this vector representation, an agglomerative clustering algorithm generates a dendrogram which results in the desired hierarchy. Starting with each pattern in a singleton cluster, clusters are merged iteratively until only one is left. The merging process is driven by linkage rules which define which two clusters are merged in each step and by the similarity measure, which is calculated between individual patterns populating a cluster. Empirical tests on the COIL-20 image set [21] using different combinations of linkage rules and similarity measures led to the conclusion that Ward's linkage rule [44] and the Tanimoto coefficient [39] as a similarity measure produce the most robust and suitable results.

3.4 Belief Update

The agent has to cope with the typical incompleteness, ambiguity and inconsistency in input data from realistic environments. We hence have developed an inference component, which uses Dempster-Shafer theory for uncertain reasoning [36]. This theory can be considered as a generalization of probability theory and distinguishes between conflicting evidence and a lack of knowledge. A basic concept of this theory is the frame of discernment Θ, which is the set of

[1] One can measure the suitability of different clustering principles by the loss of information that the resulting hierarchies introduce during the belief update.

all possible singleton hypotheses in the domain. The resulting hypothesis space 2^{Θ} comprises all possible subsets $H \subseteq \Theta$. Since we are using the hierarchical representation $\mathbf{H}$ described above, the number of relevant subsets $H \in \mathbf{H}$ is substantially reduced. The belief induced by a piece of evidence can be expressed by a mass function $m : \mathbf{H} \rightarrow [0,1]$ that assigns values to all hypotheses H such that $\sum_{H \subseteq \mathbf{H}} m(H) = 1$. A mass distribution can be interpreted as an underspecified probability distribution that preserves the unspecificity of the underlying evidence (e.g., a piece of evidence could support multiple hypotheses without committing to a specific probability distribution over these hypotheses). In addition, a belief can be equivalently described by a plausibility function pl defined as $pl(H) = \sum_{H' \cap H \neq \emptyset} m(H')$, which is sometimes more convenient.

Given all collected sensorimotor features $f_{0:t} = f_0, \ldots, f_t$, the mass distribution $m(H_t|f_{0:t})$ over the hierarchy can be recursively computed by the generalized Bayesian theorem [38,4]:

$$m(H_t|f_{0:t}) = \eta \prod_{h_t \in H_t} pl(f_{0:t}|h_t) \prod_{h_t \in H_t^C} (1 - pl(f_{0:t}|h_t)) \,, \tag{1}$$

$$pl(f_{0:t}|h_t) = pl(f_t|h_t) \sum_{H'_t \ni h_t} m(H'_t|f_{0:t-1}) \,. \tag{2}$$

Here, η is a normalization constant ensuring that the resulting mass values sum up to 1, and H_t^C denotes the complement of H_t. The plausibility $pl(f_t|h_t)$ of the new feature f_t given a hypothesis h_t (i.e., a room at the macro-level and a scene at the micro-level) is given by the relative frequency with which f_t was observed together with h_t during the training phase. Since we are using the hierarchical representation $\mathbf{H}$ instead of the full hypothesis space 2^{Θ}, the computational complexity of the update is reduced from exponential to linear [12,24]. This restriction introduces a certain error, however, due to the grouping of hypotheses sharing similar features in the clustering process, this error is largely negligible in practice.

3.5 Context Change

Compared to an earlier version, we extended the architecture by incorporating inter-room sensorimotor features in the update process using a Dempster-Shafer filter algorithm, which performs a prediction step similar to that in classical Bayes filters [15,40]. For features that indicate a state transition, the update in (1) is preceded by the following transition update. It consists of a conjunctive combination of the prior belief over H_{t-1} with the newly induced belief over H_t (under a first-order Markov assumption) by summing over all prior states H_{t-1} [5]:

$$m(H_t|f_{0:t}) = \eta \sum_{H_{t-1}} m(H_t|H_{t-1}, f_t)\, m(H_{t-1}|f_{0:t-1}) \tag{3}$$

The transition belief $m(H_t|H_{t-1}, f_t)$ can be further simplified by applying the disjunctive rule of combination $\textcircled{\cup}$ [38] in order to make the belief depend only on singletons h_{t-1} instead of aggregated states H_{t-1}:

$$m(H_t|H_{t-1}, f_t) = \bigcircledcup_{h_{t-1}\in H_{t-1}} m(H_t|h_{t-1}, f_t)\,, \tag{4}$$

$$(m_1 \circledcup m_2)\,(H) = \sum_{H_1\cup H_2=H} m_1(H_1)\, m_2(H_2)\,. \tag{5}$$

Each $m(H_t|h_{t-1}, f_t)$ is estimated in the training phase mentioned above. What is interesting here is that, in contrast to the prediction step in Bayes filters, state changes actually provide additional information which leads to a refinement of the localization belief. This is due to the fact that actions do not occur in arbitrary states but rather follow distinct sensorimotor patterns, which usually ties the act of moving in a certain way to a distinct set of locations in the environment (e.g., turning is more likely at corners).

3.6 Top-Down Uncertainty Minimization

The uncertainty minimization strategy used by the agent is based on the IBIG algorithm (*inference by information gain* [33]) and lets the agent perform actions that reduce the overall amount of uncertainty, both at the micro- and at the macro-level. Its basic principle is to determine the action a^* exhibiting the highest expected information gain with respect to the current belief distribution. The expected uncertainty is computed for each potential sensorimotor feature $\hat{f}_a = (v_1, a, v_2)$ that is compatible with the current state, i.e., v_1 and a match the current sensory input and a possible action while v_2 is integrated out. First, (1) and (3) are applied for each potential sensorimotor feature $\hat{s}_a$ corresponding to an action a in order to obtain the updated belief $m(H_t|f_{0:t}, \hat{f}_a)$. Next, the local conflict uncertainty measure I [25] is used to select the action a^* yielding the lowest expected uncertainty:

$$I(m) = \sum_H m(H) \log \frac{|H|}{m(H)}\,, \tag{6}$$

$$a^* = \arg\min_a E\left[I(m(\cdot|f_{0:t}, \hat{f}_a))\right]\,. \tag{7}$$

After executing the selected action (an eye movement at the micro-level or a change in location at the macro-level) the belief is updated with the actually observed sensorimotor feature and additional exploration steps are performed if ambiguity persists.

4 System Behavior

The following is a short description of how the components of the SMX interact in order to explore the environment and localize the agent in it. During the initial exploration phase the agents moves in the environment in order to build

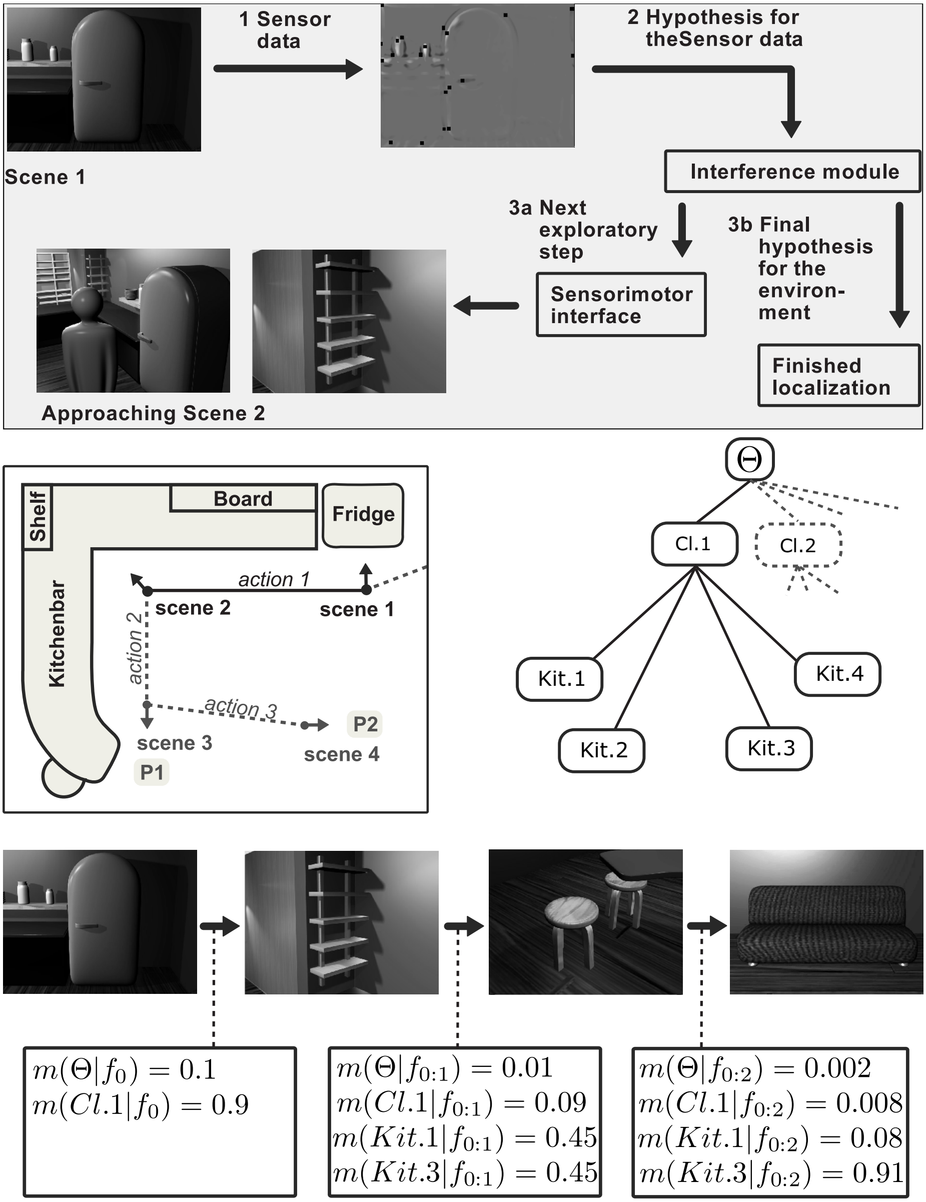


Fig. 4. Illustration of how the SMX localizes itself in the environment. An single update cycle at the macro-level is shown at the top. Parts of the environment and its hierarchical representation are shown in the middle. The lower third shows the updated beliefs after processing a sequence of sensorimotor features.

up a representation for both, the micro and macro level. Due to the way sensorimotor features are represented, salient points need to be extracted from the environment, serving as starting and destination points for actions. At the micro level, an image filter sensitive to intrinsic 2-dimensional features extracts the salient points from views, while, at the macro level, we currently use pre-defined locations in the environment. From the set of all possible sensorimotor features a large number of samples is randomly generated for both levels. Based on these samples, the hierarchical representations are build by agglomerative clustering.

The macro-level cycle starts with the agent facing a local scene. The micro-level subsystem is used to analyze this local scene by saccadic eye movements. The new sensory information returned from the micro-level exploration is a scene label which, together with the macro-level motor data and the sensory information obtained at the previous location, forms the new macro-level sensorimotor feature f_t. This feature is then used to update the current belief distribution over the hierarchy using (1) for an intra-room feature and additionally (3) for an inter-room feature (indicating a context change). Based on this, the next action is selected according to the minimum expected uncertainty as defined by (7). The agent executes this action by first rotating and then moving towards the target location. At the new location, it rotates again, if necessary, and starts a new micro-exploration by the saccadic eye movement subsystem. The cycle is repeated until a sufficient belief threshold for one of the macro-level hypotheses is reached.

A small example of a complete localization run consisting of three exploration steps is shown in Fig. 4. After processing the first sensorimotor feature, the agent has a strong belief for cluster 1, which consists of four kitchens. After processing the next feature, the evidence equally supports two of the kitchens and only the final feature completely resolves this ambiguity. A detailed quantitative analysis of the system's localization performance and of the efficiency of the action selection strategy was conducted in [51] using different virtual environments. The number of exploration steps required for sufficiently reducing localization uncertainty is considerably lower compared to the baseline of randomly performing actions (see Fig. 5), in particular for environments that exhibit high degrees of perceptual aliasing.

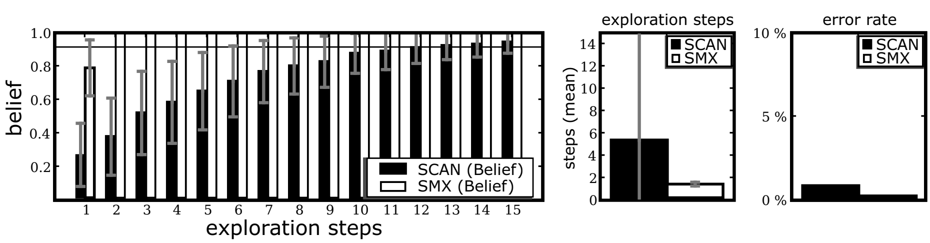


Fig. 5. Performance comparison of the IBIG exploration strategy with random action selection based on the number of required actions for reaching a belief level of 0.9 (left) and on the error rate (right). Figure adapted from [51].

5 Discussion

We developed a mobile virtual agent, SMX, which localizes itself in an indoor environment via exploratory actions at two levels of sensorimotor granularity. Compared to a previous version of the agent [51], we extended it by a belief update across different rooms in the environment and by a method for obtaining the underlying hierarchical representations in an unsupervised fashion. The former is based on a new approach for updating Dempster-Shafer belief distributions over time [27] and allows the agent to distinguish identical looking rooms and reduce localization uncertainty more quickly. Beyond that, the agent is characterized by three main properties: first, the spatial environment is represented in terms of sensorimotor features. This is motivated by doubts about the biological plausibility of map-like representations, and by psychological and neurobiological evidence suggesting a joint contribution of sensory and motor information to perception and representation. In particular, the sensorimotor representation enables the utilization of actions as an additional source of information due to their correlation with states. Second, the system operates in a loop of bottom-up processing and tow-down reasoning governed by the principle of information gain. This is achieved by a hybrid architecture in which a top-down strategy selects those exploratory actions providing the highest expected information gain with respect to the current belief. Third, the same generic hybrid architecture and information-gain strategy are used at two levels of sensorimotor granularity. This results in active localization behavior with location changes at the macro-level and saccadic eye movements at the micro-level, which mimic the way humans analyze visual scenes.

The combination of these components yields a psychologically and neurobiologically plausible system that acquires a maximum amount of information about its environments using a minimum number actions. This uncertainty minimization principle is particularly important for environments exhibiting a high degree of perceptual aliasing, e.g., rooms consisting of many similar or identical objects. The tests conducted in [51] furthermore indicate that the performance is not degraded by minor distortions of image features or of object configurations.

As mentioned, the SMX shows some differences to commonly used representations of spatial environments. The greatest difference exists with respect to those approaches that use grid-like or image-like two-dimensional maps since these do not include any explicit information about potential motor actions [6,41]. With respect to topological representations, the relation is dependent on the interpretation. If the key concept of a topological representation is seen in the abstraction from metrical properties, there is a clear difference to our approach since the motor actions in our sensorimotor representation are encoded in association with metrical information (e.g., the translation vector of an eye movement). Topological representations are also different from our approach if they are simply seen as less restricted variants of conventional spatial maps. However, it is also possible to interpret the edges in a topological graph in the sense of actions that are required to move from one node to the other, and under this perspective there is a much closer relation to the sensorimotor representation (e.g., [3,19,11,18]).

In the future development of SMX, we will apply the generic architecture to additional granularity levels of sensorimotor features. Furthermore, we will investigate the suitability of different clustering principles at these levels, e.g., by incorporating semantics of clusters [35] and by using spatial structuring principles that are better suited for problems like large-scale navigation [46,28]. This design will be guided by an ongoing evaluation based on comparisons with empirical results. On the behavioral side, this will be actual eye movements and macro-levels actions of human subjects, both in realistic and in virtual environments. On the neurobiological side, the most interesting entity for our future system development is the place cell [22]. Although there is currently no module in our system that is intended as a model of place cells, it is interesting to note that the units in our hierarchy are both influenced by the properties of the local environment and by the history of how the agent arrived at the current position, a non-trivial property that has also been observed in hippocampal neurons [8].

Acknowledgements

This work was supported by the DFG (SFB/TR 8 Spatial Cognition, project A5-[ActionSpace]).

References

1. Aloimonos, Y. (ed.): Special Issue: Purposive, Qualitative and Active Vision, Image Understanding, vol. 56 (1992)
2. Ballard, D.: Animate vision. Artificial Intelligence 48, 57–86 (1991)
3. Byun, Y.T., Kuipers, B.: A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. IEEE Journal of Robotics and Autonomous Systems 8, 47–63 (1991)
4. Delmotte, F., Smets, P.: Target identification based on the transferable belief model interpretation of Dempster-Shafer model. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 34(4), 457–471 (2004)
5. Dubois, D., Prade, H.: On the unicity of Dempster's rule of combination. International Journal of Intelligent Systems 1(2), 133–142 (1986)
6. Elfes, A.: Sonar-based real-world mapping and navigation. IEEE Journal of robotics and automation 3(3), 249–265 (1987)
7. Foo, P., Warren, W.H., Duchon, A., Tarr, M.J.: Do humans integrate routes into a cognitive map? map- versus landmark-based navigation of novel shortcuts. Journal of Experimental Psychology 31(2), 195–215 (2005)
8. Frank, L., Brown, E., Wilson, M.: Trajectory encoding in the hippocampus and entorhinal cortex. Neuron 27(1), 169–178 (2000)
9. Gadzicki, K.: Hierarchical clustering of sensorimotor features. In: Mertsching, B., Hund, M., Aziz, Z. (eds.) KI 2009. LNCS (LNAI), vol. 5803, pp. 331–338. Springer, Heidelberg (2009)
10. Gibson, J.J.: The ecological approach to visual perception. Houghton Mifflin, Boston (1979)
11. Gillner, S., Mallot, H.A.: Navigation and acquisition of spatial knowledge in a virtual maze. Journal of Cognitive Neuroscience 10(4), 445–463 (1998)

12. Gordon, J., Shortliffe, E.H.: A method for managing evidential reasoning in a hierarchical hypothesis space. Artif. Intell. 26(3), 323–357 (1985)
13. Hirtle, S.C., Jonides, J.: Evidence of hierarchies in cognitive maps. Memory and Cognition 13(3), 208–217 (1985)
14. Hommel, B., Muesseler, J., Aschersleben, G., Prinz, W.: The theory of event coding (tec): A framework for perception and action planning. Behavioral and Brain Sciences 24, 849–878 (2001)
15. Kalman, R.: A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82(1), 35–45 (1960)
16. Kohonen, T.: Self-organizing maps. Springer series in information sciences, 3rd edn., vol. 30. Springer, Heidelberg (2001)
17. Kuipers, B.: The map in the head metaphor. Environment and Behavior 14(2), 202–220 (1982)
18. Kuipers, B.: The spatial semantic hierarchy. Artificial Intelligence 119, 191–233 (2000)
19. Mataric, M.: Integration of representation into goal-driven behavior-basedrobots. IEEE Transactions on Robotics and Automation 8(3), 304–312 (1992)
20. Moore, T.: Shape representations and visual guidance of saccadic eye movements. Science 285(5435), 1914 (1999)
21. Nene, S., Nayar, S., Murase, H.: Columbia object image library (COIL-20). Tech. rep., Dept. Comput. Sci., Columbia Univ., New York (1996)
22. O'Keefe, J., Nadel, L.: The hippocampus as a cognitive map. Clarendon Press, Oxford (1978)
23. O'Regan, J.K., Noë, A.: A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences 24, 939–973 (2001)
24. Orponen, P.: Dempster's rule of combination is #P-complete. Artificial Intelligence 44(1-2), 245–253 (1990)
25. Pal, N., Bezdek, J., Hemasinha, R.: Uncertainty measures for evidential reasoning II: A new measure of total uncertainty. International Journal of Approximate Reasoning 8(1), 1–16 (1993)
26. Prinz, W.: A common coding approach to perception and action, relationships between perception and action: current approaches edn., pp. 167–203. Springer, Berlin (1990)
27. Reineking, T.: Particle filtering in the Dempster-Shafer theory. International Journal of Approximate Reasoning (2010) (in revision) (submitted) (Febuary 17, 2009)
28. Reineking, T., Kohlhagen, C., Zetzsche, C.: Efficient wayfinding in hierarchically regionalized spatial environments. In: Freksa, C. (ed.) Spatial Cognition VI. LNCS (LNAI), vol. 5248, pp. 56–70. Springer, Heidelberg (2008)
29. Reineking, T., Schult, N., Hois, J.: Combining statistical and symbolic reasoning for active scene categorization. In: Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2009, Revised Selected Papers. Communications in Computer and Information Science. Springer, Heidelberg (2010) (in press)
30. Rizzolatti, G., Craighero, L.: The mirror-neuron system 27, 169–192 (2004)
31. Rizzolatti, G., Matelli, M.: Two different streams form the dorsal visual system: anatomy and functions. Experimental Brain Research 153(2), 146–157 (2003)
32. Salton, G.: The SMART Retrieval System—Experiments in Automatic Document Processing. Prentice-Hall, Inc., Upper Saddle River (1971)
33. Schill, K.: Decision Support Systems with Adaptive Reasoning Strategies. In: Freksa, C., Jantzen, M., Valk, R. (eds.) Foundations of Computer Science. LNCS, vol. 1337, pp. 417–427. Springer, Heidelberg (1997)

34. Schill, K., Umkehrer, E., Beinlich, S., Krieger, G., Zetzsche, C.: Scene analysis with saccadic eye movements: Top-down and bottom-up modeling. Journal of Electronic Imaging 10(1), 152–160 (2001)
35. Schill, K., Zetzsche, C., Hois, J.: A belief-based architecture for scene analysis: From sensorimotor features to knowledge and ontology. Fuzzy Sets and Systems 160(10), 1507–1516 (2009)
36. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
37. Siegel, A., White, S.: The development of spatial representations of large-scale environments. Advances in child development and behavior 10, 9 (1975)
38. Smets, P.: Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. International Journal of Approximate Reasoning 9, 1–35 (1993)
39. Tanimoto, T.: IBM interal report. Tech. rep., IBM (November 1957)
40. Thrun, S., Fox, D., Burgard, W., Dellaert, F.: Robust Monte Carlo localization for mobile robots. Artificial Intelligence 128(1-2), 99–141 (2001)
41. Thrun, S.: Learning occupancy grids with forward sensor models. Autonomous Robots 15, 111–127 (2003)
42. Tversky, B.: Distortions in cognitive maps. Geoforum 23(2), 131–138 (1992)
43. Wang, R.F., Spelke, E.S.: Updating egocentric representations in human navigation. Cognition 77, 215–250 (2000)
44. Ward, J.H.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58(301), 236–244 (1963)
45. Wiener, J., Mallot, H.: 'Fine-to-coarse' route planning and navigation in regionalized environments. Spatial Cognition and Computation 3(4), 331–358 (2003)
46. Wiener, J., Mallot, H.: 'Fine-to-coarse' route planning and navigation in regionalized environments. Spatial Cognition and Computation 3(4), 331–358 (2003)
47. Yarbus, A.L.: Eye Movements and Vision. Plenum Press, New York (1967)
48. Zetzsche, C., Galbraith, C., Wolter, J., Schill, K.: Navigation based on a sensorimotor representation: a virtual reality study. In: Rogowitz, B.E., Pappas, T.N., Daly, S.J. (eds.) Proceedings of SPIE. Human Vision and Electronic Imaging XII, February 2007, vol. 6492 (2007)
49. Zetzsche, C., Krieger, G.: Nonlinear operators and higher-order statistics in image processing and analysis. In: Proc. ISPA 2001 - 2nd International Symposium on Image and Signal Processing and Analysis, pp. 119–124 (2001)
50. Zetzsche, C., Wolter, J., Galbraith, C., Schill, K.: Representation of space: image-like or sensorimotor. Spatial Vision 22(5), 409–424 (2009)
51. Zetzsche, C., Wolter, J., Schill, K.: Sensorimotor representation and knowledge-based reasoning for spatial exploration and localisation. Cognitive Processing 9, 283–297 (2008)

Color Binding in Visuo-Spatial Working Memory

Luca Simione[1,2,3], Antonino Raffone[1,2,3], Gisella Micciantuono[1,2], Marta Olivetti Belardinelli[1,2], and Cees van Leeuwen[3]

[1] Department of Psychology, "Sapienza" University of Rome, Italy
[2] ECONA (Interuniversity Center for Cognitive Processing in Natural and Artificial Systems), Rome, Italy
[3] Laboratory for Perceptual Dynamics, BSI RIKEN, Japan
luca.simione@gmail.com

Abstract. Visuo-spatial working memory (VSWM) is characterized by a limited storage capacity. An important issue of vision and memory research is whether VSWM stores integrated objects or features. Following Luck and Vogel' (1997), we investigated VSWM for color-color conjunctions. We used a single probe at test to minimize configuration-related effects on performance, and two baselines with storage of single-feature (color) objects. We found that performance with color conjunctions (binding) was even lower than in the condition in which the same amount of features was presented at different object locations. Neural network simulations show that our experimental evidence can be transparently explained in terms of local processing limitations, co-existing with global processing and VSWM storage limitations. Behavioral predictions are put forth.

Keywords: Visuo-spatial working memory, binding, neural networks, storage capacity, processing capacity.

1 Introduction

Visuo-spatial working memory (VSWM) plays a crucial role in the active mental representation of the visual world, to serve various aspects of cognitive processing (Baddeley and Hitch, 1974; Bundesen, 1990). A striking functional property of VSWM is given by its limited capacity (Sperling, 1960; Phillips, 1974; Bundesen, 1990; Luck and Vogel, 1997; Raffone and Wolters, 2001).

It is crucial to estimate such storage capacity limitations without contamination from other storage modalities in working memory, and finding out the units of representation in such a store. Thus, Luck and Vogel (1997) proposed a new experimental paradigm to assess VSWM capacity, by introducing a VSWM task based on change detection (see also Phillips, 1974). In this paradigm, a memory array and a test array with a variable number of objects, are separated by a short time interval (about one second). In half of the trials, the memory and the test arrays are identical; in the other half there is a change of a target feature, such as color or orientation, of one of the items. In this change detection paradigm the sample display is presented for a short time period (100 ms), thus minimizing the possibility for the items to be verbally encoded.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 179–190, 2010.

With this experimental paradigm, Luck and Vogel (1997; see also Vogel, Woodman et al. 2001) estimated a capacity of visual working memory of about four objects by applying Pashler's formula (Pashler, 1988; see Section 2.2.) on the accuracy data. Interestingly, Luck and Vogel (1997) found that in an experiment of their series in which subjects had to store color-orientation conjunctions, performance was not different from conditions in which orientation only or color only where dimensions relevant for change detection. Related to this evidence, in their last experiment they assessed storage of color-color conjunctions, with a small colored square presented within a square frame with a different color. Remarkably, color change detection performance in such a color-color conjunction condition was not different from performance in the conditions in which either the small colored squares or the larger colored square frames were presented. Thus, Luck and Vogel concluded that VSWM capacity does not depend on the number of features making up the objects, with integrated objects rather than individual features being the units of representation in such working memory store. This is the so-called "strong object hypothesis" (Olson and Jiang, 2002), for which VSWM is limited only for number of objects that can be stored and not for number of features stored for each object.

Subsequent studies with the same paradigm however led to alternative suggestions about binding in VSWM, including the possibility of completely independent storage systems for each feature dimension (e.g. color, spatial orientation, size; Wheeler and Treisman, 2002). Following the "weak object hypothesis" (Olson & Jiang, 2002), the VSWM capacity is determined both by the number of objects and the number of features, but features from different domains can be bound together to form an object. This hypothesis supposes that change detection performance for isolated features is poorer that performance for conjoined objects, given the same number of features in both kinds of displays. Thus, the weak object hypothesis suggests a facilitation in object storage in comparison with single feature storage. In contrast with both these two hypotheses, Schneider et al. (2001) found an impairment in change detection performance by presenting objects composed of two conjoined colors, in comparison with conditions in which single colored squares were presented (remarkably, they uses the same experimental setting as in Luck and Vogel's study in 1997). Thus, contrary to a facilitation, these authors found an interference effect in the color conjunction condition.

Given this controversial theoretical and empirical scenario, the aim of the present study was to provide a further assessment of whether storage of color-color conjunctions (binding) in VSWM is consistent with the strong object hypothesis, the alternative weak object hypothesis, or a further alternative interference-based account. We use the same experimental paradigm and setting as in Luck and Vogel's (1997) study, except for presenting a single probe at test, to prevent any confounding configuration-related effect. Moreover, to contrast in a fine fashion color binding-related performance with single-color performance, we assessed color-color conjunction performance against top and bottom baselines, in which either half or the same amount of features in single-feature (color) objects were presented, respectively (see Section 2.1). Finally, to shed light on underlying neurocognitive processes mechanisms, we developed and studied an explicit connectionist model, after the experiment, incorporating three levels of representation and processing the visual system. We will first present the experiment, then the model, and finally an integrated discussion of our findings.

2 Color Binding Experiment

2.1 Methods

Participants.
Twelve psychology students from the “Sapienza” University of Rome participated in the study. All participants (5 male, mean age 24.2, in the range 21-30), had a normal or corrected-to-normal visual acuity, normal color discrimination, and no history of neurological problems. They were naïve as to the purpose of the study, which lasted for approximately 40 minutes.

Stimuli.
The stimuli were presented on a computer screen of 15.4’’ (refresh rate = 60 Hz) located in a quiet and dark room. The distance between the participant’s head and the video monitor was approximately 60 cm. Three types of stimuli were used: small squares (0.65° × 0.65°), large squares (1.3° × 1.3°) or small colored squares inside a larger colored square (see Figure 1). The color of each object was randomly selected (with replacement) from a set of four colors: blue, green, red and purple, with the restriction that if a small square appeared in the same space location of a large square these colors were different from each other.

Stimuli presentation was programmed with E-Prime 1.0 software.

Design and Procedure.
Each trial began with the presentation of a fixation cross on the center of the screen for 2000 ms. Then the memory array appeared for 100 ms, containing a variable number of colored squares. The squares were either all large or all small (with a respective randomized occurrence in 50% of the trials) in Top and Bottom baseline conditions; in the color binding condition a small square appeared inside a larger square frame of a different color, at each object location. After a retention interval of 900 ms, one of the objects previously displayed in the memory array appeared at the same location. This object (probe) could appear with the same color as in the memory array (in 50% of the trials, with a random occurrence), or in a different color. In the color binding condition, either the small square or the large square frame might change color (in 25% of the trials, respectively).

As in Luck and Vogel’s (1997) study, the participants were instructed to report whether a color change occurred by matching memory and test arrays, by key pressing (the keys “A” and “K” were used for response, counterbalanced across subjects). Three blocks (one for each condition) of 60 trials (20 for each of the three set sizes) were run, for a total of 180 trials. Each block was preceded by 12 training trials. The set sizes were different for each block: for the Top baseline, the set sizes were 2, 3 and 4 items; for the Bottom baseline the set sizes were 4, 6 and 8 items; for the Conjunction condition the set size were “2+2”, “3+3” and “4+4” items (thus comprising the two kinds of colored squares conjoint at object locations. i.e. the amount of features in the array was displayed in half the number of objects).

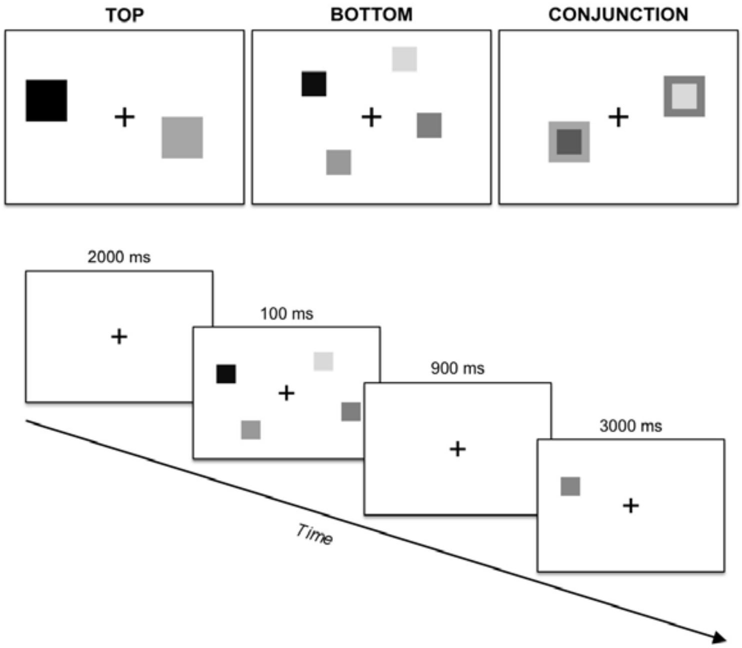


Fig. 1. Stimulus examples (on the upper panel) and experimental procedure (on the bottom panel). Note that different levels of the grayscale are used to represent different colors and that proportion between stimuli and display size is not indicative of real proportion. See Methods for more details.

2.2 Results

A two-way ANOVA on the accuracy data with the within-participants factors of Condition (Top baseline, Bottom baseline and Conjunction) and Set Size (Min., Middle and Max.), was conducted. We also conducted planned comparisons between the conjunction condition and the other two conditions (Top baseline vs. Conjunction condition, and Bottom baseline vs. Conjunction condition). The percentage accuracy results of the experiment (means and standard errors) are displayed in Figure 2.

The ANOVA analysis revealed a significant main effect of Condition [$F(2) = 47.5$ $p < 0.001$] and Set Size [$F(2) = 20.5$ $p < 0.001$] on accuracy performance, but no significant interaction between these two factors [$p > 0.05$]. Planned comparisons between different conditions revealed that accuracy in the Conjunction condition (Min. = 78%; Middle = 66%; Max. = 62%) was lower than in the Top baseline (Min. = 92%; Middle = 87%; Max. = 85%) for all the set sizes [$p < 0.001$]. Interestingly, the same was observed for the Conjunction versus Bottom baseline conditions (Min. = 85%; Middle = 77%; Max. = 72%) for all set sizes [$p < 0.05$].

Additionally, we computed the mean number of objects encoded (stored) in each condition by using the Cowan's k formula: for each set size, $k = N$ x (hit rate + correct rejection rate - 1), where k is the number of encoded objects or estimated storage capacity, and N is the number of objects presented. We thus computed an average over the set sizes in the bottom baseline and in the Conjunction condition (in the top baseline the set sizes were half the ones in the other two conditions). We found a k = 3.38 for the Bottom baseline and k = 2.5 for the Conjunction condition. Thus, we found that the estimated memory capacity in the Bottom baseline was of about one object higher than in the Conjunction condition.

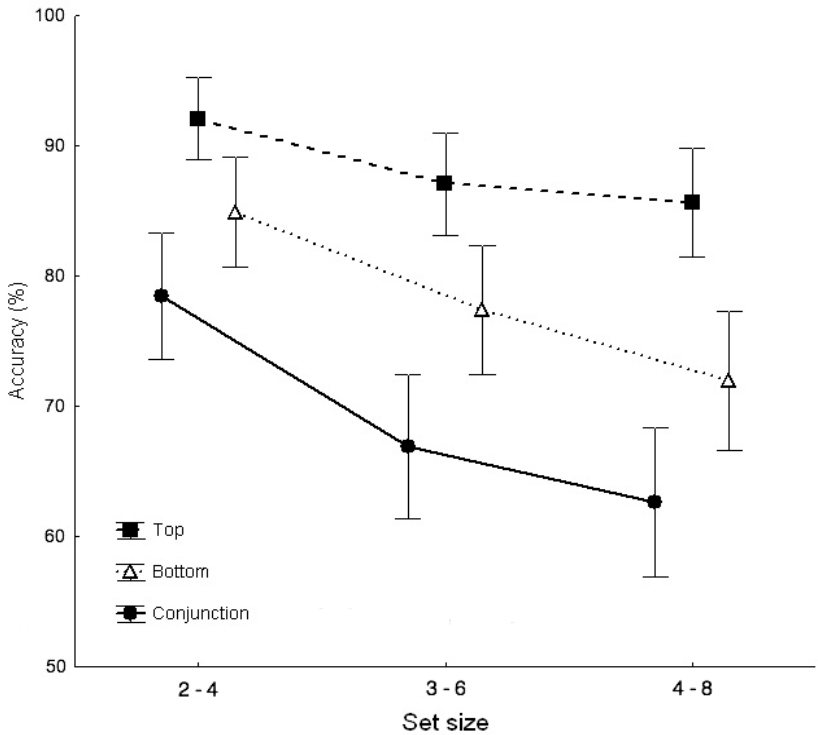


Fig. 2. Percentage of accuracy for factor of Condition (Top, Bottom and Conjunction) in function of set size in the first experiment. Error bars represent the standard error of the means.

3 Color Binding Model

Computational models have proven to be very useful to shed light on the underlying mechanisms for limited storage capacity in visual working memory (Raffone & Wolters, 2001; Johnson et al., 2009). Raffone and Wolters' (2001) model accounted for Luck and Vogel's results by a neural synchrony mechanism, consistent with the strong object hypothesis. However, this model did not consider the case of color-color conjunctions and the contrasting evidence related to it. Moreover, it did not consider evidence and functional aspects consistent with the weak object hypothesis.

Our present computational model is devoted to account for our present experimental results about color binding for storage in VSWM. As Raffone and Wolters (2001), and unlike Johnson et al. (2009), we did not explicitly simulate the change detection (retrieval) process, but only the neural processes related to encoding and maintenance. Our model incorporates simple and transparent mutual inhibition based mechanisms for both limited storage capacity of VSWM, and for a limited perceptual processing capacity (see Bundesen, 1990; Bundesen et al., 2005), both in terms of local and global processing competition. It is crucial to estimate such storage capacity limitations without contamination from other storage.

3.1 Neural Units

Neural units in the model are rise/decay integrators described by an activity level x. The evolution of x follows Equation 1:

$$\frac{dx}{dt}\tau = -x + (1-x)\left(\sum exc + I_{noise} + I_{input}\right) - \left(x\sum inh\right) \tag{1}$$

where Σ_{exc} and Σ_{inh} are the sums of excitatory (including self-excitatory) and inhibitory inputs, respectively, I_{noise} is a Gaussian noise input ($\mu = 0.05$, s.d.= 0.4), randomly extracted at each time step, I_{input} is the direct visual input (with strength $\gamma_{input} = 0.8$) given to the units in the first layer during stimulus presentation, and τ is a time constant setting the rise ('gain') and decay rate of units' activity.

The excitatory and inhibitory inputs in most cases (except when indicated) are passed through a sigmoid function f(x) (see also Grossberg & Grunewald, 1997), with given Q and n parameters, as in Equation 2:

$$f(x) = \frac{x^n}{Q^n + x^n} \tag{2}$$

3.2 Model Architecture

The model includes three layers or simulated brain areas, inspired by recent neuroimaging evidence about brain correlates of visual working memory (Todd & Marois, 2004). We also took into account realistic connection delays in inter-area connection (see Dehaene et al., 2003).

The three layers of the model correspond to different cortical regions: a V1-V3 sensory layer, mostly involved in ongoing perceptual processing; an intermediate layer standing for area V4 and the lateral occipital complex (V4-LOC) area, shown to be involved in object feature processing for storage in visual working memory (see Xu & Chun, 2006); a VSWM layer, corresponding to parietal posterior (PP) cortex.

In each area there is a neural unit encoding for each object presented to the network. For the sake of simplicity, we only simulate forward connectivity between the three layers, with a one-to-one connectivity for each object. In the PP layer, self-excitation enables the self-sustained delay activity necessary for VSWM maintenance. The time constant τ of neural unit integration (decay) is longer for the intermediate V4-LOC layer to simulate an iconic memory stage (τ = 2 ms for V1-V3 and PP layers, and τ = 16 ms for V4-LOC). Connection strengths are denoted with γ, and signaling delays with δ.

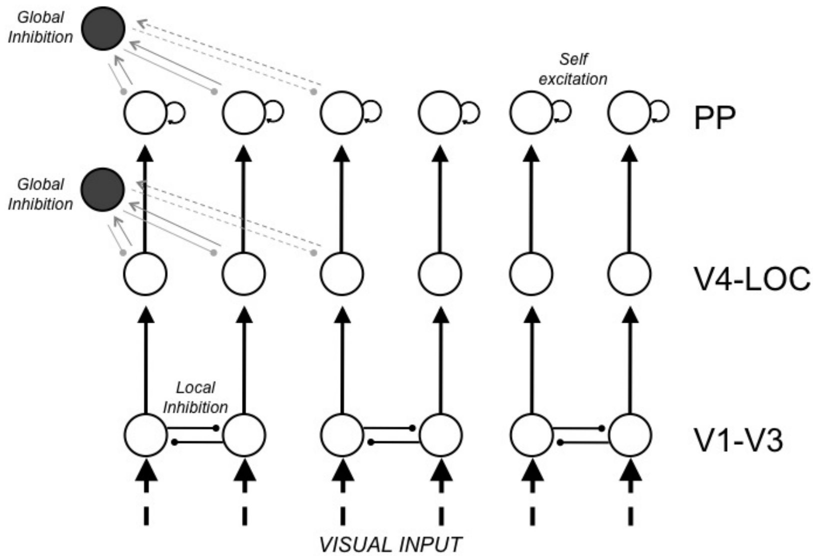


Fig. 3. Model architecture. A neural unit (empty circle) is present in each area for each object. Filled circles represents global inhibitory units, that is connected with all the neural unit in the same area, receiving excitatory connections and sending inhibitory connections to each one. Arrow pointers represent excitatory connections (or input, as in the first area) whereas circle pointers represent inhibitory connections.

Mutual inhibition plays a crucial role in the model, for different forms of inter-object competition in the three layers: it acts "locally" in a linear fashion in the V1-V3 layer, i.e. only between neural units encoding for adjacent colored square in a color-color conjunction; it acts at a global competition level in a linear fashion to "attenuate" graded activations in the intermediate iconic V4-LOC layer; it acts at a global level in a non-linear fashion to "suppress" a subset of non-linear activations in the VSWM PP layer (k-winners-take all dynamics, where k is the storage capacity of VSWM).

The 'external' visual inputs arrive to the V1-V3 layer. Mutual inhibition in this layer is linear, with $\gamma = 1.3$ and $delay = 2$ ms. V1-V3 units send forward outputs ($\gamma = 1.1$; $delay = 6$ ms; $Q = 0.4$ and $n = 2$) to V4-LOC. In this layer mutual inhibition is weak, so object- related activity is only attenuated. Units in V4-LOC have a longer time constant, thus tenths of milliseconds are necessary for the slower rise of their activity level. When activity in a V4-LOC unit approaches a certain level, object-related activity is transferred via forward connections ($\gamma = 0.2$; $delay = 6$ ms; $Q = 0.2$ and $n = 4$) to the PP layer for VSWM storage.

In the PP layer activity is self-sustained by self-excitatory connections ($\gamma = 1.5$; $delay = 2$ ms; $Q = 0.3$ and $n = 6$). There is a strong global non-linear mutual inhibition between PP units ($\gamma = 0.8$; $delay = 2$ ms; $Q = 1.8$ and $n = 25$), with a k-winners-take-all dynamics in which only three to four objects can be maintained at the time by their corresponding units. In the mutual suppressive interactions in the PP layer the activity of other units does not reach the level of self-excitation for maintenance. Thus in our

VSWM model object representations are consolidated in a all-or-non manner, according to recent experimental evidence (see Zhang & Luck, 2008).

3.3 Simulation Results

The aim of our simulations is to account for our present experimental results on storage in VSWM of color-color conjunctions, by keeping into account local processing, global processing and VSWM storage capacity limitations. In the simulations, a millisecond of real time corresponds to 5 simulation cycles.

In each simulation, direct visual input was given after 200 ms of startup for 100 ms. Then there were other 1700 ms of maintenance during with no other stimuli. Thus each simulation lasted for a total of 2000 ms. Each reported simulation result is obtained averaging thirty simulation trials, unless it is specified otherwise.

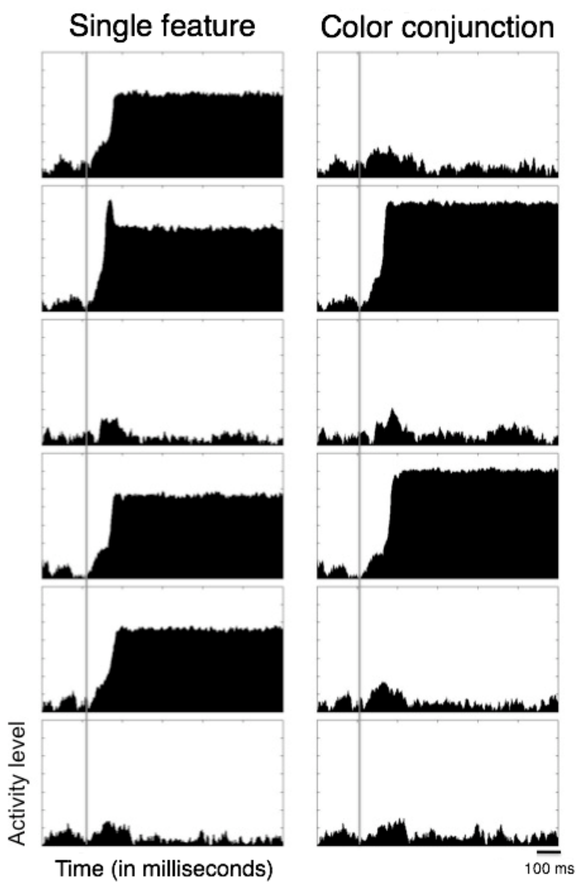


Fig. 4. Activity for each neural unit in Posterior Parietal area for single feature (left panels) and color binding (right panels) conditions. As shown, in the single feature conditions there are four active neural units, whereas in the color binding condition only two. The vertical gray line indicates the time at which the stimulus arrives.

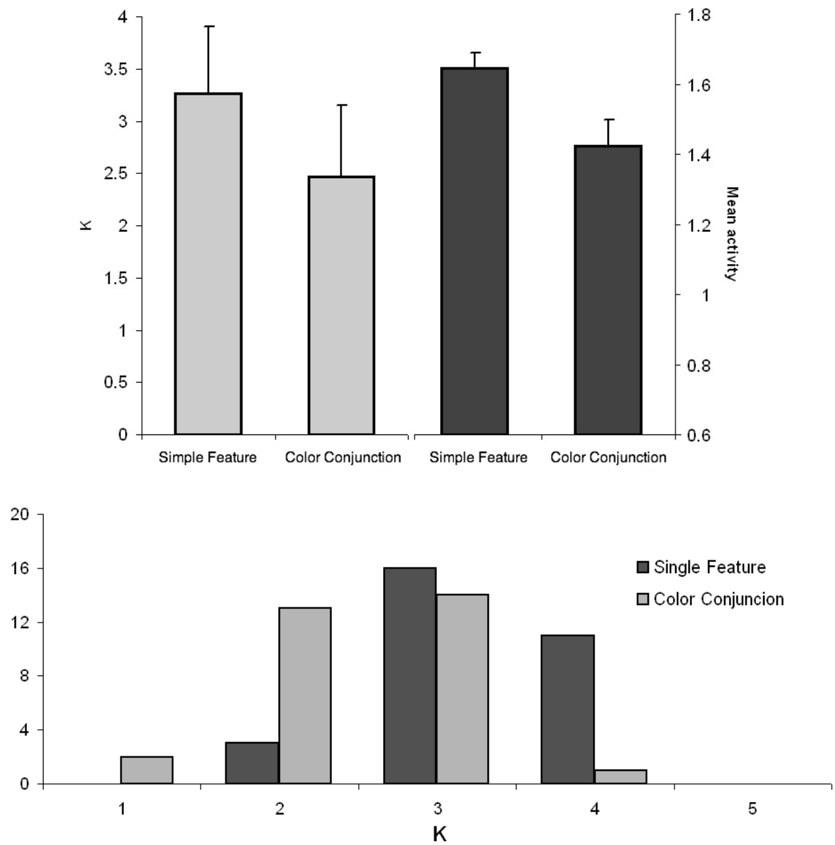


Fig. 5. Simulation results for single feature and color binding conditions with short stimuli presentation. In the upper panel, average numbers of stored objects (K) in light gray and average PP area activity in dark gray. Error bars represent the standard deviation. In the bottom panel, distribution of K for single feature (dark gray) and color binding (light gray) conditions.

Figure 4 displays a typical PP activity in two different simulated conditions with six presented objects. In the first condition, named "single feature" condition, each object is defined by only one feature and thus there is no (local) mutual inhibition in V1-V3. In the second condition, named "color conjunction" (or color binding) condition, each object is defined by two conjoined features, thus in the V1-V3 layer a local mutual inhibition is established between these two features. As a result, V1-V3 activity in the color conjunction condition is reduced with respect to the single feature condition, implying more time for V4-LOC units to reach the activity level at which they in turn start to spread activation to the associated PP units. Thus, as soon as the visual input is removed and the activity in V1-V3 and V4-LOC falls down, in the PP layer a maintenance-related difference between the two simulated conditions is observed: in the single feature condition the PP storage capacity is fully employed (up to four objects), whereas in the color conjunction condition only two objects are stored in this simulation trial, with the same parameters, initial values and noise values in the two simulation conditions.

As shown in Figure 5, we accounted for our experimental findings for the single feature and the color conjunction conditions, with a drop in the number of objects maintained in VSWM in the color conjunction condition with respect to the single feature condition. We calculated the number of maintained objects by simply counting the number of units in which the mean activity level, from stimulus onset to the end of the trial, was higher than a threshold value (θ = 0.3) at which maintenance versus non-maintenance was clearly observed. We thus refer to the number of maintained objects in our model as to *k* to match this with our experimentally-estimated *k*. In our model thus the local mutual inhibition in V1-V3 results in a drop of about 1 object in the color conjunction condition, as shown in the experimental color conjunction condition. Interestingly, the same drop was found in the mean layer activity,

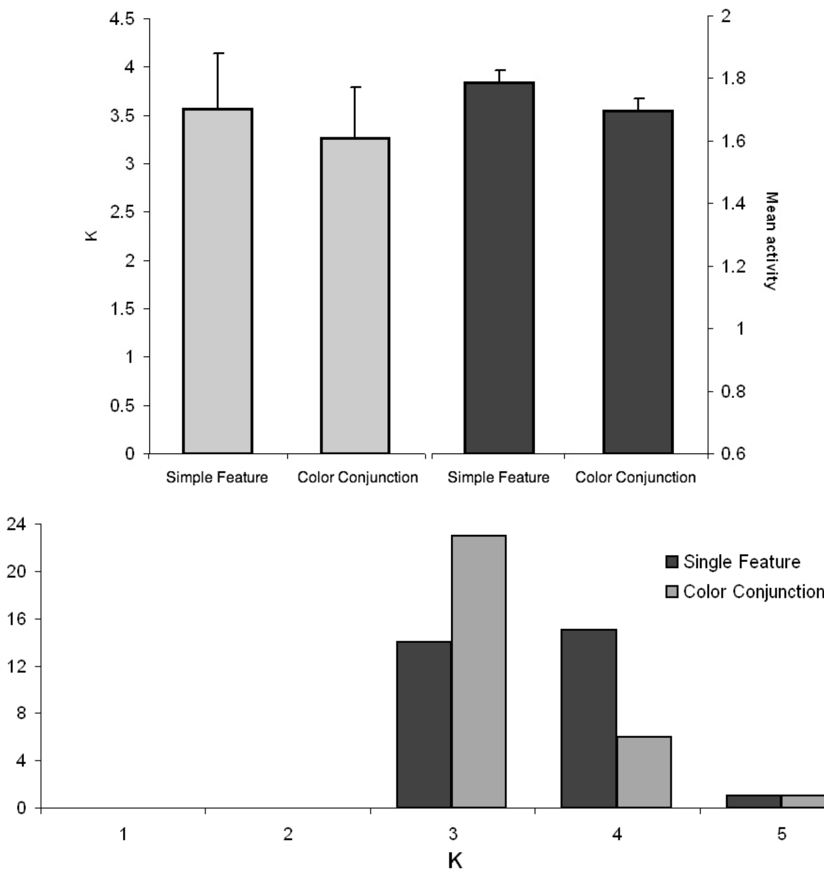


Fig. 6. Simulation results for single feature and color binding conditions with long stimuli presentation. In the upper panel, average numbers of stored objects (*K*) in light gray and average PP area activity in dark gray. Error bars represent the standard deviation. In the bottom panel, distribution of *K* for single feature (dark gray) and color binding (light gray) conditions.

which might be linked to a brain region activity index, computed over all the units, for the entire simulation duration and all the simulation trials for the same conditions. The last panel of Figure 5 shows the distribution of k values in the thirty simulations run for each condition. The most frequent value is $k = 3$ for both the conditions, but for the single feature condition we have a higher occurrence of $k = 4$ than for $k = 2$, whereas for the color conjunction condition the reverse pattern was observed.

However, it hardly likely that VSWM capacity 'contracts' with color conjunctions. As the functional logic of our neural network model suggests, it seems plausible that local competitive interactions take place between adjacent colors and reduce the flow of activation at iconic level feeding the VSWM store, which might be related to the construct of limited processing capacity (Bundesen, 1990). In this case, a dissipation of perceptual processing capacity would occur due to local competitive interactions. However, it can be hypothesized with that with a longer exposure time of the memory array would counteract this dissipation.

We thus performed a simulation with a longer stimulus exposure (500 ms versus 100 ms). The simulation results are displayed in Figure 6: both the mean k and the mean activity level differences between the two simulated conditions are greatly reduced with a longer exposure time. In the same manner, the k distribution of frequencies shows no differences between the single feature condition and the color conjunction condition, with most occurrences distributed between $k = 3$ and $k = 4$.

4 Discussion

Our experimental findings about storage of color-color conjunctions in VSWM are consistent with a local processing competition account rather than with a strong object hypothesis or a weak object hypothesis. How can differences in results by using the same experimental procedure be explained? Unlike Luck and Vogel (1997), we used a single probe at test, thus reducing potential configuration-related effects on performance. Moreover, also in light of Schneider et al.'s (2001) findings, differences between the participant samples might also have contributed to the observed differences in results.

Related to this aspect, it can be hypothesized that practice leads to a color chunking counteracting and even dominating local color interference. Thus, 'chunking fields' and synchronizing neural connections implemented in Raffone and Wolters' (2001), associated to the strong object hypothesis and the original Luck and Vogel's (1997) study, and the local inhibition implemented in the present model, associated to a local processing competition account, might jointly be implemented in a further computational model, after a further practice-focused experimental investigation. Thus, the "chunking field" hypothesis might represent a synthesis beyond the weak and strong object hypothesis-antithesis. Long-term memory would thus play an important role in binding for visual working memory, as also suggested by Baddeley (2000).

A further VSWM experiment with a longer exposure time would provide a direct test of the local competition account emerging from our present experiment and associated computational investigation. Finally, an ERP (event-related potential) experiment would capture the different latencies in memory array processing with an without color-conjunctions, with the same displays used in our present experiment (and in Luck and Vogel's study), as related to our neural network model. Different ERP waves for memory array processing and VSWM maintenance (Vogel & Machizawa, 2004), depending on the experimental condition, are expected.

References

1. Baddeley, A., Hitch, G.: Working memory. In: Bower, G.H. (ed.) The psychology of learning and motivation: Advances in research and theory, vol. 8. Academic Press, New York (1974)
2. Bundesen, C.: A theory of visual attention. Psychol. Rev. 97(4), 523–547 (1990)
3. Bundesen, C., Habekost, T., Kyllingsbaek, S.: A neural theory of visual attention: bridging cognition and neurophysiology. Psychol. Rev. 112(2), 291–328 (2005)
4. Johnson, J.S., Spencer, J.P., Schöner, G.: A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. Brain Res. 1299, 17–32 (2009)
5. Luck, S.J., Vogel, E.K.: The capacity of visual working memory for features and conjunctions. Nature 390(6657), 279–281 (1997)
6. Olson, I.R., Jiang, Y.: Is visual short-term memory object based? Rejection of the "strong-object". Percept Psychophys 64(7), 1055–1067 (2002)
7. Pashler, H.: Familiarity and visual change detection. Perception & Psychophysics 44, 369–378 (1988)
8. Phillips, W.A.: On the distinction between Sensory storage and short-term visual memory (1974)
9. Raffone, A., Wolters, G.: A cortical mechanism for binding in visual working memory. J. Cogn. Neurosci. 13(6), 766–785 (2001)
10. Schneider, W.X., Deubel, H., Wesenick, M.B.: Characterizing chunks in visual short-term memory: Not more than one feature per dimension? Behav. Brain Sci. 24(1), 144–145 (2000)
11. Vogel, E.K., Woodman, G.F., Luck, S.J.: Storage of features, conjunctions and objects in visual working memory. J. Exp. Psychol. Hum. Percept. Perform. 27(1), 92–114 (2001)
12. Wheeler, M.E., Treisman, A.M.: Binding in short-term visual memory. J. Exp. Psychol. Gen. 131(1), 48–64 (2002)
13. Zhang, W., Luck, S.J.: Discrete fixed-resolution representations in visual working memory. Nature 453(7192), 233–235 (2008)
14. Dehaene, S., Sergent, C., Changeaux, J.P.: A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proc. Natl. Acad. Sci. U.S.A 100(14), 8520–8525 (2003)
15. Grossberg, S., Grunewald, A.: Cortical synchronization and perceptual framing. Journal of Cognitive Neuroscience 9(1), 117–132 (1997)
16. Todd, J.J., Marois, R.: Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428(6984), 751–754 (2004)
17. Xu, Y., Chun, M.M.: Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440(7080), 91–95 (2006)
18. Sperling, G.: The information available in brief visual presentations. Psychol. Monogr. 74 (1960)
19. Baddeley, A.: The episodic buffer: a new component of working memory? Trends Cogn. Sci. 4(11), 417–423 (2000)
20. Vogel, E.K., Machizawa, M.G.: Neural activity predicts individual differences in visual working memory capacity. Nature 428(6984), 748–751 (2004)

Human EEG Correlates of Spatial Navigation within Egocentric and Allocentric Reference Frames

Markus Plank[1], Hermann J. Müller[2], Julie Onton[3], Scott Makeig[3], and Klaus Gramann[3]

[1] Institute for Neural Computation, University of California San Diego, 9500 Gilman Drive #0523, La Jolla, CA 92093, USA
[2] Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München, Germany
[3] Swartz Center for Computational Neuroscience, University of California San Diego, 9500 Gilman Drive #0559, La Jolla, CA 92093, USA
mplank@ucsd.edu

Abstract. We investigated the impact of path complexity on brain dynamics of subjects who preferentially use an egocentric (Turners) or an allocentric (Nonturners) reference frame during spatial navigation. Participants indicated a return bearing direction ('point-to-origin') following visual presentation of virtual tunnel passages, varying with respect to the complexity of the outbound path. High-density electroencephalographic activity was recorded continuously and spatially filtered with Independent Component Analysis. For Turners, rotations and translations were associated with decreased and increased (8-12 Hz) alpha activity in occipito-parietal cortex, whereas Nonturners displayed increased alpha within cortical areas along the ventral pathway, as well as in retrosplenial cortex, an area supporting bidirectional exchange of information between parietal and medial temporal regions. Both groups displayed complexity-related modulations of frontal midline (4-8 Hz) theta activity. Findings extend results of hemodynamic imaging and neuropsychological studies on spatial navigation and emphasize the need for considering individual proclivities when investigating human navigation performance.

Keywords: path integration, reference frames, egocentric, allocentric, EEG, ICA, Independent Component Analysis, source reconstruction, ERSP.

1 Introduction

Spatial navigation constitutes a sub-category of spatial cognition and denotes the capacity to plan and execute goal-directed paths based on the computation, maintenance, and utilization of internal representations of the environment [1]. These representations comprise the location of threats, rewards, and other agents and their spatial relations, as well as one's own position with respect to the represented entities [2]. The navigator's position and orientation can be inferred by *path integration*, i.e., the continuous integration of local translations and rotations from movement cues [3].

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 191–206, 2010.

The integration of spatial information during path integration takes place within dissociable, but interacting spatial reference frames. Most generally, a distinction is made between a self-centered *egocentric* reference frame and an environment-centered *allocentric* reference frame [2, 4]. Within the former, *egocentric* self-to-object *distances* and *bearings* are coded independently of the global layout of the environment, with respect to the three intrinsically defined axes of the navigator (front–back, right–left, and up–down). As the navigator moves, egocentric parameters have to be updated with each consecutive step by adding the displacement vector of an object (relative to the navigator) to their previous egocentric position vectors. In other words, the world constantly changes, whereas the navigator remains spatially fixed in the center of the reference system [5]. The resulting spatial representation therefore can be characterized as being highly dynamic and transient. Behavioral studies in triangle-completion have indeed shown that with increasing path complexity in terms of overall length and/or number of turns, spatial updating of egocentric parameters becomes more and more challenging, resulting in larger errors and increased response times [6, 7].

By contrast, an *allocentric reference frame* establishes a coordinate system with an origin and a reference direction external to the navigator. Within the resulting *allocentric locational representation* inter-object relations are represented independent of the navigator's current position and/or orientation, exclusively related to the external reference properties, with coordinate axes being defined by the global layout of the environment. Also, the navigator himself is represented in terms of position, but without any orientation. The use of an allocentric reference frame requires the moving navigator to constantly update his position but not his orientation since all *allocentric distances* and *allocentric bearings* remain stationary as the navigator proceeds. Recent experiments have provided initial evidence for spatial accuracy being considerably high even after long and winding outbound trajectories whenever participants represent the traversed pathway within an allocentric reference frame [8].

The subject's choice of navigating within an egocentric or an allocentric reference frame partly depends on the sensory modality activated during navigation [9], as well as the perspective from which an environment is initially encountered and learned [10, 11]. Further, several studies have provided rich evidence for the existence of intraindividually stable proclivities in using either an egocentric or an allocentric frame during visual path integration [12-14]. By developing a task that allowed for a categorization of subjects with respect to their preference for using an egocentric or an allocentric reference frame [12], it was possible to differentiate *Turners*, who preferentially use an egocentric reference frame, and *Nonturners*, who preferentially use an allocentric reference frame for spatial navigation (see Figure 1). This segregation was neither attributable to differences in the feeling of vection nor to group-specific eye movement patterns during navigation through sparse visual environments [15].

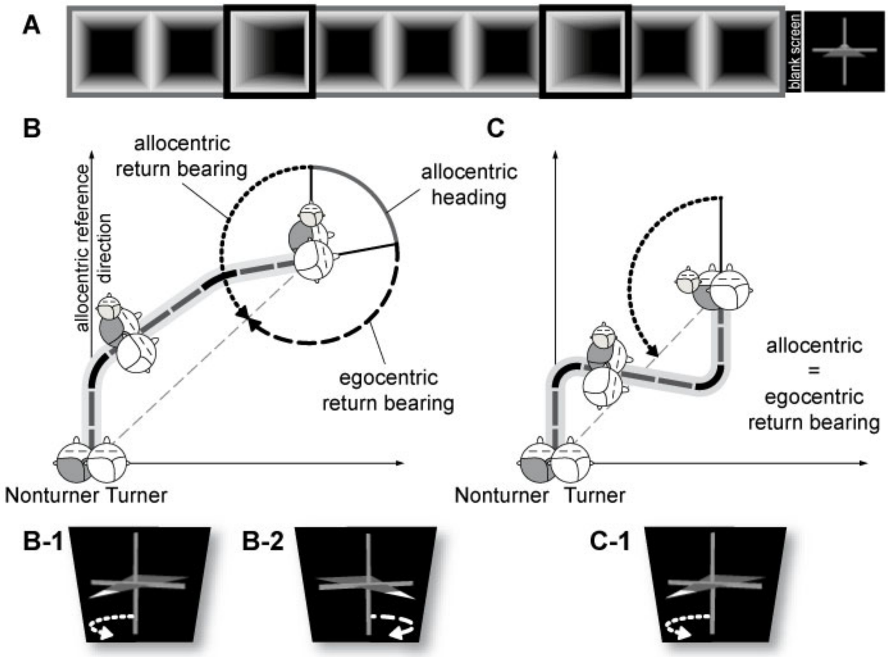


Fig. 1. **[A]** Snapshot of a tunnel with straight and curved segments (framed black), providing visual information on forward motion from a first person perspective; **[B]** When traversing tunnels with two turns bending to the same side, return bearings of Turners (using an egocentric reference frame, white heads) and Nonturners (using an allocentric reference frame, dark grey heads) are initially aligned, but diverge with changes in perceptual heading along the trajectory. At the end, Nonturners rotate the virtual arrow so that it points 120° to their left (**B-1**), whereas Turners adjust the arrow 120° to their right (**B-2**). The homing responses indicate that Nonturners adjust the homing arrow with respect to the allocentric reference direction (small light grey head) corresponding to the allocentric return bearing, whereas Turners respond with respect to their final cognitive heading corresponding to the egocentric return bearing; **[C]** For tunnels with two opposite turns of equal angularity, both systems are re-aligned after the second turn, resulting in identical arrow adjustments for Turners and Nonturners (**C-1**), based on the equivalence of egocentric and allocentric return bearings.

1.1 Cortical Structures and Processes Subserving the Use of Egocentric and Allocentric Reference Frames

The conceptual dissociation of egocentric and allocentric reference frames is further reflected by structural as well as functional differences in the underlying neural substrate [11, 16]. Under natural circumstances, both systems interact during the encoding and retrieval of spatial knowledge. Based on incoming visual flow, information regarding one's ego-motion is extracted and further processed along the *dorsal pathway* within medial occipital and posterior parietal areas [17], with the latter coordinating transformations between various reference frames (e.g., eye-, body-, head- or world-centered) [18, 19]. Firing patterns of parietal cells are affected by combinations of velocity, head direction, and visual stimuli, bearing resemblance to findings on parietal lesions in humans associated with improper integration of heading information [20, 21].]

Additional cells that respond selectively to position-independent heading are located in cingulate cortex [22]. Particularly its posterior portion (retrosplenial cortex) is interconnected with posterior parietal cortex, superior temporal sulcus, as well as subcortical structures [23, 24]. The retrosplenial cortex plays a central role for the proper integration of head direction and information on heading-invariant landmark information arising from place cells in medial temporal, particularly hippocampal structures. Place cells in the hippocampal formation have been found to respond to an animal's location in space independent of its current orientation [25], giving rise to the assumption that certain aspects of the allocentric reference frame reside in these areas [26]. Beyond object-based processing along the *ventral pathway*, hippocampal structures also receive context-independent position information upstream from entorhinal grid cells. In contrast to place cells, these cells establish a map of the environment that is only initially anchored to external landmarks, but persists in their absence, which might contribute central elements for the long-term storage of spatial structures [27]. Finally, parietal as well as hippocampal structures possess interconnections to prefrontal regions that are associated with reference-frame unspecific functions of spatial working memory comprising maintenance of encoded information as well as goal-directed planning [28].

The temporal dynamics of activity within these areas during the acquisition, consolidation, and retrieval of spatial knowledge may be investigated by means of high-density electroencephalography (EEG). Generally, spontaneous EEG activity is constituted by uncoupled intracortical sources producing random oscillatory activity in a wide frequency range. Sensory stimulation causes a coupling of these generators, resulting in temporally synchronous and coherent oscillations. In the context of spatial navigation, alpha and theta frequency bands constitute the most extensively studied oscillations, since they have repeatedly been shown to correlate with mental states and strategies as well as stimulus characteristics. Activity in the (4 – 8 Hz) theta frequency band is directly related to memory maintenance and increases with task difficulty [29]. Particularly during heading changes pronounced theta activity has been detected [30] which might serve as gating mechanism during information encoding, consolidation and retrieval [31]. By contrast, activity in the (8 – 13 Hz) alpha frequency band has several functional correlates reflecting sensory, motor, and memory functions [32]. Alpha arises primarily from posterior sites, including occipital, parietal, and posterior temporal regions. Decreased alpha constitutes a valid signature of activation or cognitive preparedness of the cortical domain for processing of task-related information [33].

However, the EEG signal measured on the scalp surface does not originate on-site, i.e., directly beneath a certain electrode, but can be characterized as a sum of many electrical processes, including those with neural or muscular origin [34, 35]. The far-field potential arising from each of these synchronized cellular assemblies spreads via volume conduction and is recorded, to a greater or lesser extent, by every surface electrode instantaneously. Independent Component Analysis (ICA) has been successfully applied to EEG data to separate these mixed signals into spatially fixed but temporally independent processes [36].

1.2 Aims of the Current Study

Bringing together the excellent temporal resolution of high-density EEG recordings, seminal progress in data mining techniques, as well as VR technology, the current study aimed at the identification of generator sources of brain activity as well as macroscopic oscillatory dynamics during path integration within egocentric and allocentric reference frames. The difference in primitive parameters between egocentric and allocentric spatial representations suggests that they are processed within distinct neural circuits (dorsal vs. ventral pathway). On the other hand, following the remarks of Burgess [4] as well as initial results of Gramann et al. [13], egocentric and allocentric systems might be activated in parallel. This parallel processing might be reflected by specific dynamics within distinct brain areas responsible for the processing of reference frame-specific spatial information, as well as reference frame-unspecific activation associated with the coordinated transformation of spatial information between reference frames. Based on this question, the systematic variation of the path layout allowed, for the first time, the analysis of how brain dynamics within areas associated with reference frame-specific as well as -unspecific cognitive processing are modulated by changes in outbound path complexity.

2 Material and Methods

Thirty-seven male students recruited from the Ludwig-Maximilians-University Munich, Germany, participated in the Experiment (age (M±SD) = 24.64 ± 3.52 years). Participants were either paid 8€ per hour or received course credit for taking part in the experiment. All subjects had normal or corrected to normal vision and reported no history of neurological disorder.

Prior to the main experiment, subjects were classified for their preferred spatial strategy (Nonturner vs. Turner). In this categorization task, participants sat in an electromagnetically shielded room, 170 cm in front of a flat projection screen (120 x 90 cm), which was illuminated by a Sanyo PLC-XU47 projector (800 x 600 pixels, 275W, 60 Hz). The screen center was horizontally aligned with the participant's line of sight. Visual stimulation covered a visual field of view of approx. 41° x 41°. Participants passively traversed 30 tunnels with one single turn of varying angle. Tunnels were composed of an initial straight segment followed by a curved segment and two final straight segments. Movement was simulated solely by optic flow, with a constant speed of 2.27 seconds/segment. At the end of a passage two virtual arrows were displayed simultaneously that represented the correct homing response based on an allocentric or an egocentric reference frame. In a forced-choice task, participants had to spontaneously decide which of the two arrows pointed back to the starting point of their passage (see Figure 1B for examples of homing arrows).

Within three blocks of 10 trials each turning angles gradually decreased, so that egocentric and allocentric arrows converged, resulting in increasing task difficulty. In order to take part in the main experiment, participants had to consistently (i.e., more than 83% of the trials) select the allocentric or egocentric arrow to be categorized as Nonturners or Turners, respectively. All subjects selected for the main experiment demonstrated consistent choice of one or the other reference frame (Turners M±SD =

99% ± 0.02%), Nonturners M±SD = 96% ± 0.04%; correlation between preferred strategy and strategy-specific arrow choice: $r(37) = .997$, $p < .00001$). 20 subjects were categorized as *Turners* (egocentric) and 17 as *Nonturners* (allocentric). All Turners were right-handed, four Nonturners were left-handed.

The main experimental session included 20 blocks of 9 tunnels each with minor in-between breaks. During the experiment tunnels with one turn and two turns were presented. The task of the subjects was to maintain orientation while being passively transported along the tunnel passage and to subsequently rotate a simulated 3-D arrow to point directly back to the starting point of the trajectory ('point-to-origin'). In order to minimize EEG movement artifacts, arrow adjustment was accomplished by pressing and holding right and left mouse buttons. Pressing the center mouse button confirmed the adjustment.

Tunnels with one turn consisted of five segments (segment duration = 1875 ms), with the two initial and two final straight segments enclosing the turning segment of varying angle. Tunnels with two turns had a total of nine segments (turn segments located in segments 3 and 7). Since depth of sight was limited to the upcoming 1.5 segments, the placement of the first turn in segment 3 ensured that participants could not determine the direction and amount of the first rotation directly at the beginning of the presentation. Further, we inserted three straight segments between the first and the second turn, so that subjects were not able to predict whether they were traversing a tunnel with five or with nine segments until either the end of the passage appeared or the tunnel continued through a second turn. In 10% of the trials participants received strategy-specific feedback on homing accuracy (based on the a priori categorized preference).

The main experiment was composed of 180 trials, constituting a factorial combination of 2 lengths/number of turns (5 segments, one turn; 9 segments, two turns) × 2 sides of end position (left; right) × 4 categorical eccentricities of end position[1] (15°, 30°, 45°, and 60°, each with variation of ± 2° in order to prevent stereotypical homing responses) × 10 repetitions per condition, plus 20 filler trials (start of each block) being straight tunnels.

The EEG was continuously recorded with a high-density array of 128 Ag/AgCl electrodes corresponding to the international 5%-system [37]. Impedance was kept below 7 kΩ. The signal was filtered online with a band-pass of .016–100 Hz, and digitized at a sampling rate of 500 Hz. An additional electrode was placed on the infraorbital ridge of the left eye to record the vertical electro-oculogram (EOG). Data were analyzed off-line with the freely available MATLAB-toolbox EEGLAB [38]. After downsampling to 250 Hz data were digitally filtered to remove frequencies above 50 Hz and re-referenced to linked mastoids. Continuous data were first screened for atypical (excessive peak-to-peak deflections or bursts of electromyographic activity) epochs. The remaining data were decomposed by extended infomax ICA using binica [39] as implemented in EEGLAB.

[1] Eccentricity of end position was computed within an allocentric reference frame (with respect to the initial straight segment of the tunnel passage). Different end positions corresponded to different turning angles, e.g., the end position of 15° eccentricity required either a single turn of 30°, or two turns to the same side of 15° and 10°, or two opposite turns of 30° and -30°.

For every participant, ICA returned 128 maximally independent components. Subsequently, DIPFIT2 routines from EEGLAB were applied to fit single equivalent dipole models to the IC scalp topographies using a four-shell spherical head model [40]. Components with bilaterally distributed scalp maps were fit with a dual dipole model using a symmetrical constraint. Components with an equivalent dipole model explaining less than 85% variance of the measured scalp maps and components that reflected muscle activity, electrocardiogram, or eye movements were excluded from further analysis.

After decomposition, the data were segmented into overlapping epochs of 2875 ms (length of one tunnel segment plus 500 ms pre- and 500 ms post-stimulus time window) and normalized for between subject comparison by subtracting mean log power from single-trial log power. Subsequently, data were baseline-corrected by subtracting the average EEG spectrum of all trials, beginning with the second tunnel segment and ending with the penultimate segment of the tunnel passage (for tunnels with one turn: Segment 4; for tunnels with two turns: Segment 8). Mean Event-Related Spectral Perturbations (ERSP) were computed by subtracting the 2-D (frequency-by-latency) mean log power spectrum of the baseline from the mean log power at each time point of the experimental trials [41, 42]. This approach provides insights into event-related brain dynamics in a wide frequency range for the whole time course of the tunnel passage.

IC processes were clustered across subjects by means of a joint distance measure, based on power spectra, event-related potentials, scalp projections, equivalent dipole locations, mean ERSP, and inter-trial coherence for each selected IC from each subject. The resulting component distances were clustered using a *K-means* cluster algorithm as implemented in the EEGLAB toolbox (see [34] for further details).

3 Results and Discussion

Strategy-specific expected and observed angular adjustments were highly correlated for both Nonturners and Turners. For Turners, using an egocentric reference frame, correlation coefficients were comparable at all complexity levels ($r_{1\text{-turn}}(80) = .950$; $r_{2\text{-turns(same)}}(70) = .924$; $r_{2\text{-turns(opposite)}}(60) = .899$; all ps < .0001). This pattern was identical for Nonturners, using an allocentric reference frame ($r_{1\text{-turn}}(120) = .972$; $r_{2\text{-turns(same)}}(72) = .937$; $r_{2\text{-turns(opposite)}}(48) = .\ 895$; all ps < .0001). Although tunnels varied on a trial-to-trial basis with respect to the turning angle and overall length, the pronounced correlation between expected and observed homing responses confirms the subjects' capability to extract directional information from purely visual flow on local translations and rotations. Given that the underlying egocentric and allocentric representations of Turners and Nonturners were of comparable spatial accuracy, it was of interest, if the strategy groups differed in terms of the underlying cortical structures and spectral dynamics within these areas during navigation.

A total of 599 independent component (IC) processes were retained (see Figure 2); 275 from Nonturner subjects, and 324 from Turner subjects. The number of ICs ranged from 9 to 28, with both groups displaying a mean number of retained components of 16.2 (difference between strategy groups: $\chi^2_{df=1} = 1.05$, $p > .05$). 22 IC clusters were obtained. Two clusters mirrored horizontal and vertical EOG activity, and

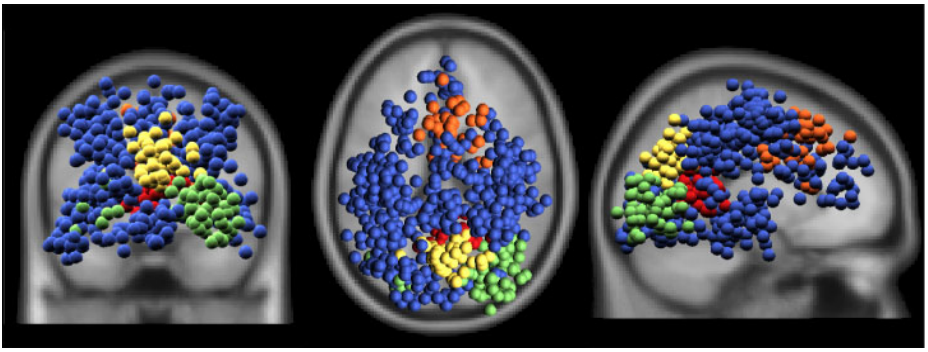


Fig. 2. Reconstructed equivalent model dipole locations of the 599 functional IC processes (dipoles related to eye movements are not shown); Blue spheres display equivalent model dipole locations of IC processes in 16 clusters without significant ERSP differences between strategy groups and/or complexity levels; (green) cluster located in or near bilateral occipital gyrus; (yellow) cluster in or near right precuneus; (red) cluster in or near posterior cingulate/retrosplenial cortex; (orange) cluster in or near right cingulate gyrus

20 clusters reflected brain processes (see Figure 2). Stereotaxic Talairach coordinates, residual variances, Brodmann Areas, the number of Nonturner and Turner subjects, as well as the number of ICs within four representative clusters are summarized in Table 1. The results replicated and extended previous findings regarding the spectral dynamics of distinct brain areas during spatial navigation [13, 43] by further analyzing the brain dynamics dependent on the complexity of the outbound path. At all complexity levels,

Table 1. Properties of 4 representative IC clusters, sorted from posterior to anterior IC cluster sites (along the y-axis). Columns provide information regarding **(1)** the location of the cluster centroids in Talairach space (x-y-z). All reconstructed clusters for each condition were anatomically specified within the stereotaxic coordinate system of Talairach and Tournoux using the Talairach demon software [45], returning the coordinates of the nearest grey-matter point. Further, the table provides information regarding **(2)** the residual variance (RV, in %) of the reconstructed cluster centroids, and **(3)** their anatomical region defined in the Brodmann Area system [46]. Finally, the table gives information regarding the number of Nonturner and Turner subjects **(S_{NT}, S_T)**, as well as the amount of Nonturner and Turner Independent Components **(IC_{NT}, IC_T)** within each cluster.

Cl	Talairach			RV	Brodmann Areas		S_{NT}	S_T	IC_{NT}	IC_T
	x	y	z	[%]						
1	29	-79	-1	3.41	BA 18	R (bilat.) inferior occipital gyrus	13	17	22	33
2	4	-69	30	2.65	BA 7/31	R (midline) precuneus	13	15	16	23
3	5	-49	11	5.37	BA 29/30	R posterior cingulate/ retrosplenium	9	9	16	17
4	7	13	38	3.56	BA 32	R cingulate gyrus	12	10	20	14

Turners and Nonturners displayed comparable deflections from baseline in or near primary/secondary visual cortex (Independent Component Cluster ICC 1, BA 17/18) reflecting activity accompanying visual processing from a first-person perspective. Whereas translational information was associated with relative synchronization of the alpha frequency band (near 10 Hz), the combined translational/rotational information during tunnel turns resulted in alpha desynchronization (see Figure 3-A). The desynchronization of alpha has been associated with increased cognitive processing during relevant time-points of a given task [31, 44].

Despite identical visual processing of Turners and Nonturners, spectral dynamics displayed distinct patterns as information was processed along the dorsal stream, particularly in or near posterior parietal cortex (ICC 2, BA 7/31). Whenever the second turn bended into the same direction, Turners exhibited a relative alpha desynchronization in or near posterior parietal cortex upon viewing the second turn prevailing until they entered the turning segment. By contrast, Nonturners' alpha activity did not differ from baseline during this time period (see Figure 4-A). In posterior parietal cortex the fusion of multiple egocentric reference frames is accomplished, further projecting to motor structures in order to coordinate appropriate movements [47]. In concert with retrosplenial cortex, posterior parietal cortex further integrates heading changes [48], conveying allocentric aspects of space [49]. Since the homing response of Turners closely resembles the egocentric return bearing that requires the updating of cognitive heading during rotations, the more pronounced alpha desynchronization of Turners points to increased cognitive effort during the integration of future heading during turns [50, 51]. The homing response of Nonturners, by contrast, matches the allocentric return bearing that, by definition, does not require the integration of heading. The absence of alpha desynchronizations during turns, however, does per se not exclude the integration of egocentric information for Nonturners. Rather, the processing of this information is not as accentuated during turns to the same side as for Turners.

For tunnel configurations with two *opposite* turns of equal angularity (Figure 4-B), Turners and Nonturners displayed significant alpha desynchronization in posterior parietal regions upon viewing the second turn. But whereas for Nonturners alpha power increased to baseline as they actually entered the turn, for Turners alpha desynchronization persisted until the end of the second turn.

Given that the general side of egocentric self-to-object relations (including the egocentric return bearing) can be maintained when traversing subsequent turns to the same side but that opposite turns necessitate the whole environmental array to be shifted with respect to the intrinsic axis of the navigator, the prolonged processing of rotational information for Turners might be associated with a cognitively highly demanding spatial updating of egocentric return bearing as well as heading during the second, opposite turn [3]. Our results indicate that this shift of relative self-to-object relations is accomplished in posterior parietal cortex, in line with previous studies [e.g., 51].

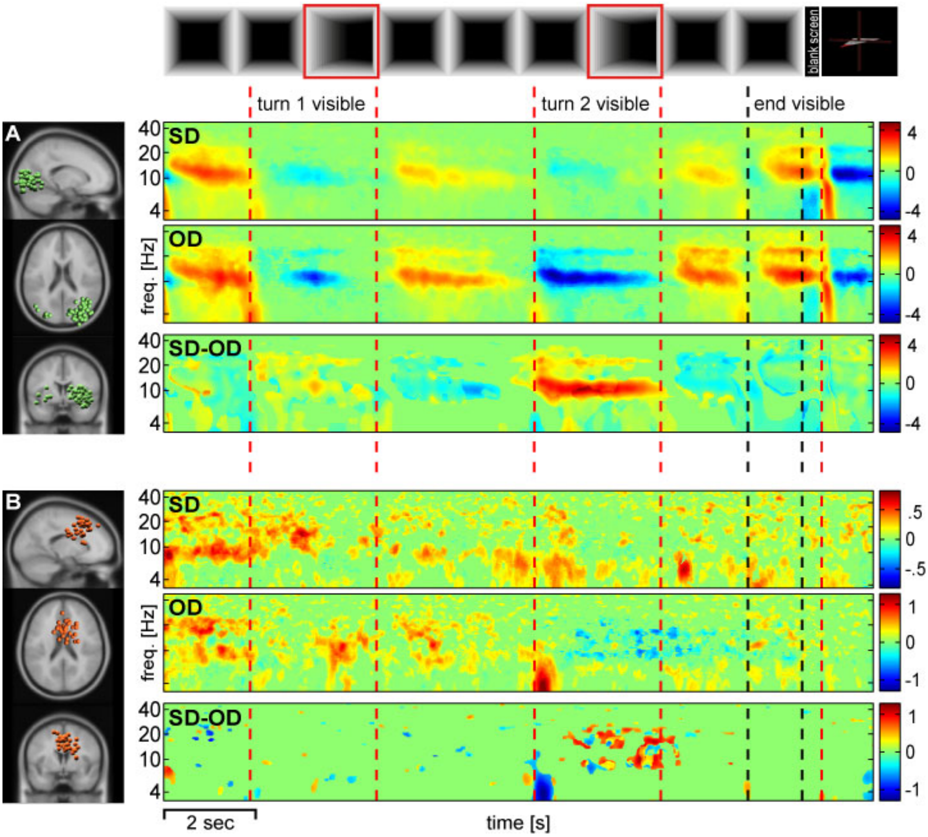


Fig. 3. [A] Mean ERSP images for a cluster of independent source processes located in or near bilateral lingual gyrus (BA 17/18) revealing task-dependent changes in spectral power during spatial navigation through tunnel passages containing two turns into the same direction (**SD**) and into opposite directions **(OD)**. Cluster centroid mean ERSPs are plotted in log-spaced frequencies from 3 – 45 Hz. Green indicates no significant ($p > .001$) difference in mean log power (dB) from baseline. Warmer colors indicate significant increases (synchronizations, dB > 0), and colder colors decreases (desynchronization, dB < 0) in log power from baseline. ERSP differences between tunnel configurations are shown in panel **(SD-OD)** ($p < .0001$, corrected for multiple comparisons); Important time points of the tunnel passage are marked with dashed lines, indicating the period when participants approached and traversed turns, or encountered the end of the tunnel, as well as the initial 1.38 seconds of the virtual homing vector adjustment; **[B]** ERSP images for a cluster of IC processes located in or near right cingulate gyrus (BA 32), separately for tunnels with two turns into the same direction **(SD)**, and two turns into opposite directions **(OD)**. Panel **(SD-OD)** depicts ERSP differences between tunnel configurations.

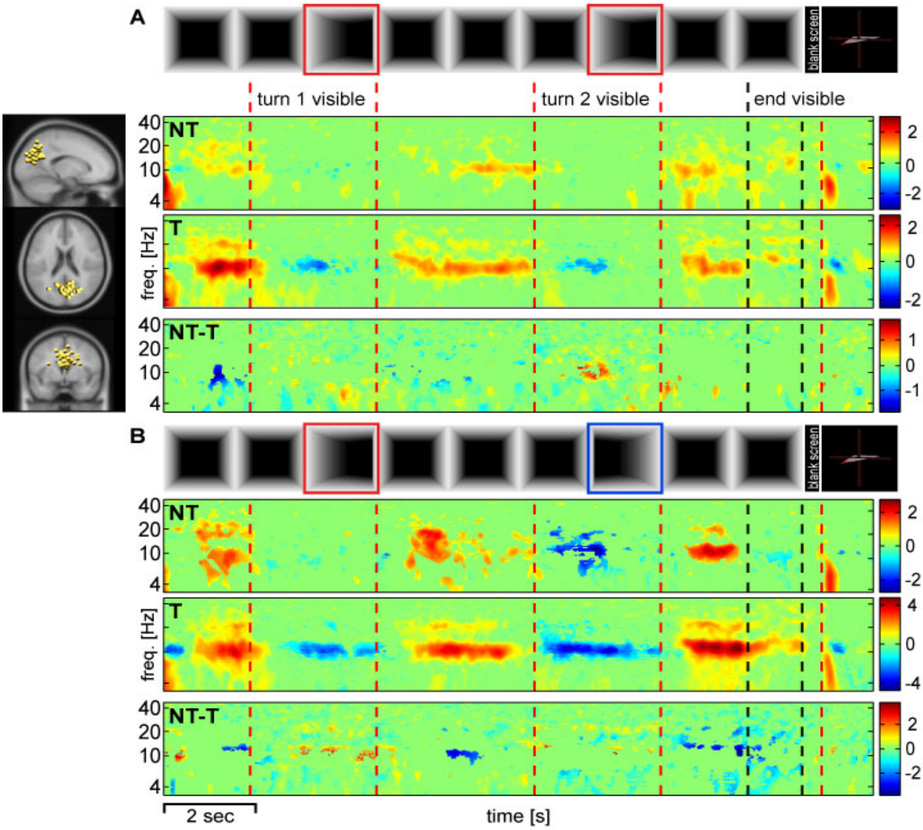


Fig. 4. Mean event-related spectral perturbation (ERSP) images for an IC cluster located in or near midline precuneus (BA 7/31) revealing task-dependent changes in spectral power during spatial navigation through tunnel passages containing two turns into the same direction **[A]** and into different directions **[B]**, with separate plots for Nonturners, navigating within an allocentric reference frame **(NT)**, and Turners, using an egocentric reference frame **(T)**. ERSP differences between strategy groups are shown in panels **(NT-T)** ($p < .0001$, corrected for multiple comparisons). For further explanation see Figure 2.

As can be seen from Figure 5, prior to entering as well as during the second, opposite turn, only Nonturners demonstrated significantly increased alpha desynchronization in or near retrosplenial cortex (ICC 3, BA 29/30). Since actual perceptual heading (egocentric) and cognitive heading (allocentric) diverge during this time period, this finding might be associated with a more pronounced inhibition of processing egocentric visuospatial information in order to prevent interference with maintaining two spatial reference frames in parallel [52]. However, it remains an open question why this pattern is only prevalent during the second, opposite turn. In addition to the strategy-specific spectral dynamics of the (near 10 Hz) alpha frequency band in occipital and parietal cortices, subjects displayed comparable theta (near 4 – 8 Hz) synchronizations in or near medial frontal cortex that occurred at specific time-points of the trajectory, e.g., as turns or the end of the passage became visible (see Figure 3-B).

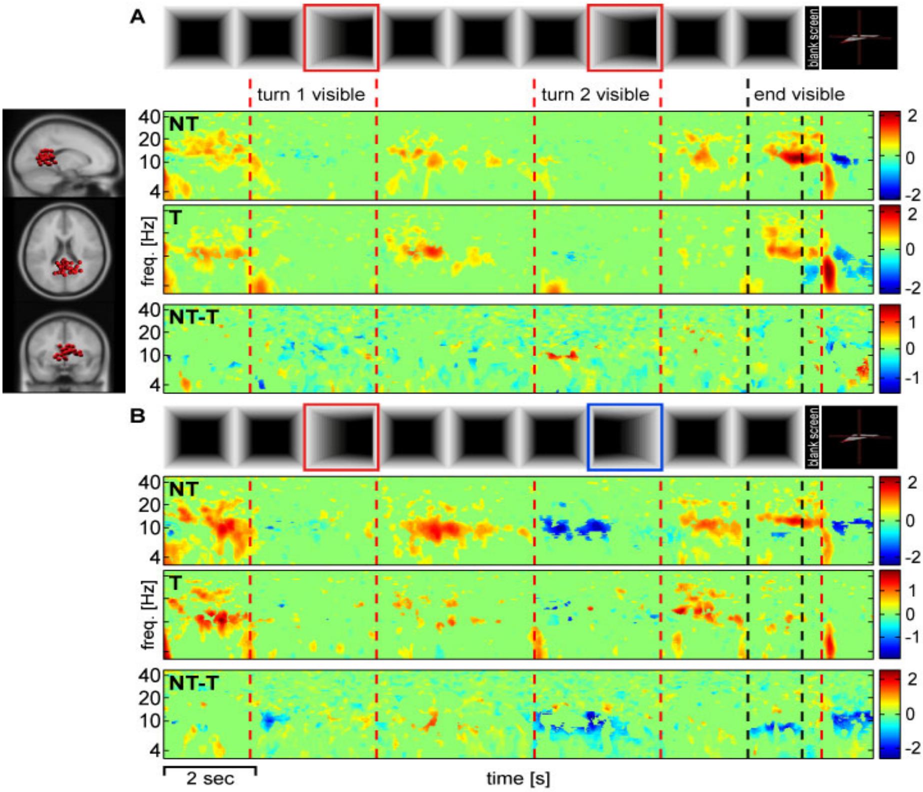


Fig. 5. Mean event-related spectral perturbation (ERSP) images for an IC cluster located in or near midline posterior cingulate cortex (retrosplenium, BA 29/30) revealing task-dependent changes in spectral power during spatial navigation through tunnel passages containing two turns into the same direction **[A]** and into different directions **[B]**, with separate plots for Non-turners, navigating within an allocentric reference frame **(NT)**, and Turners, using an egocentric reference frame **(T)**. ERSP differences between strategy groups are shown in panels **(NT-T)** ($p < .0001$, corrected for multiple comparisons). For further explanation see Figure 2.

Importantly, for tunnels with two turns the actual time-point of the final theta synchronization in or near anterior cingulate cortex (ICC 4, BA 32) was determined by the layout of the outbound path: In case of two turns bending into the same direction, the final theta synchronization appeared not until after the final turn, resembling the pattern for tunnels with one turn (not shown). For tunnels with two *opposite* turns of equal angularity, it already emerged upon viewing the second turn. Since turns of opposite tunnel configurations always were of equal angularity, participants might have determined the angle of the upcoming second turn even before entering the turn itself. This could have prevented the occurrence of a final theta burst anteceding the final turning segment. By contrast, whenever the second turn bent into the *same direction*, angles were not equal but several path layouts could end on a certain eccentricity. Here, participants first had to process the rotational information during the second turn before reliably updating their heading and return bearing [30]. Therefore, this activity pattern in frontal theta might index cognitive effort during spatial updating based on the retrieval of the previously traversed tunnel segments [53, 54].

4 Conclusion and Perspective

To the author's knowledge, this is the first study investigating EEG brain dynamics accompanying the use of an allocentric or egocentric reference frame during path integration on multiple complexity levels. The individual preference for the use of an egocentric or an allocentric reference frame was accompanied by dissociable brain dynamics in distinct cortical networks. Turners and Nonturners displayed differential spectral activation patterns of the alpha frequency band (near 8 – 12 Hz) within several IC clusters in or near occipital, occipito-parietal as well as posterior parietal cortices. Additional differences were found in or near retrosplenial cortex. These differences might be linked to a strategy-specific updating of representational primitives such as egocentric and allocentric return bearings and distances as well as allocentric heading from integration of rotational and/or translational information. Besides strategy-specific modulations, Turners and Nonturners displayed comparable, complexity-modulated activation patterns of the theta frequency band (near 4 – 8 Hz) in or near medial frontal cortex most likely mirroring increased cognitive effort during memory retrieval of previously traversed segments.

Taken together, the results suggest that the complexity of the traversed pathway in terms of the direction of subsequent turns has differential effects on macroscopic brain dynamics of subjects preferring to navigate within an egocentric or an allocentric reference frame. Since identical visual stimulation was found to provoke differential activity in brain areas linked to reference-specific cognitive processes, individual proclivities should be considered more carefully in studies on spatial navigation. Future research has to address how the present results transfer to more general navigation tasks, e.g., when additional vestibular and proprioceptive information or landmarks are available. Fully immersive 3-D VR systems utilizing mobile brain imaging [55] might constitute the seminal basis for further investigation of complexity effects on behavioral and electrocortical correlates of allocentric and egocentric navigation under more natural conditions.

Acknowledgements. This research was supported by the German Research Foundation (DFG GR2627/2-1), the German Academic Exchange Service (DAAD), the Boehringer-Ingelheim-Fonds, and by the G.-A.-Lienert-Foundation for Promotion of Young Researchers in Biopsychological Methods. The authors further want to thank the anonymous reviewers for their help in improving this article, as well as the participants for their time and interest in this work.

References

1. Gallistel, C.R.: The organization of learning. MIT Press, Cambridge (1993)
2. Klatzky, R.L.: Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In: Freksa, C., Habel, C., Wender, K.F. (eds.) Spatial Cognition 1998. LNCS (LNAI), vol. 1404, pp. 1–17. Springer, Heidelberg (1998)

3. Loomis, J.M., Klatzky, R.L., Golledge, R.G., Philbeck, J.W.: Human navigation by path integration. In: Golledge, R.G. (ed.) Wayfinding behavior. Cognitive Mapping and Other Spatial Processes, pp. 125–151. John Hopkins Univ. Press, Baltimore (1999)
4. Burgess, N.: Spatial memory: How egocentric and allocentric combine. Trends in Cognitive Sciences 10, 551–557 (2006)
5. Wang, R.F., Spelke, E.: Human spatial representation: Insights from animals. Trends in Cognitive Sciences 6, 376 (2002)
6. Loomis, J.M., Klatzky, R.L., Golledge, R.G., Cicinelli, J.G., Pellegrino, J.W., Fry, P.A.: Nonvisual navigation by blind and sighted: Assessment of path integration ability. Journal of Experimental Psychology - General 122, 73–91 (1993)
7. Klatzky, R.L., Loomis, J.M., Golledge, R.G., Cicinelli, J.G., Doherty, S., Pellegrino, J.W.: Acquisition of route and survey knowledge in the absence of vision. Journal of Motor Behavior 22, 19–43 (1990)
8. McNamara, T.P., Rump, B., Werner, S.: Egocentric and geocentric frames of reference in memory of large-scale space. Psychonomic Bulletin & Review 10, 589–595 (2003)
9. Avraamides, M.N., Klatzky, R.L., Loomis, J.M., Golledge, R.G.: Use of cognitive versus perceptual heading during imagined locomotion depends on the response mode. Psychological Science 15, 403–408 (2004)
10. Montello, D.R., Waller, D., Hegarty, M., Richardson, A.E.: Spatial memory of real environments, virtual environments, and maps. In: Allen, G.L. (ed.) Human Spatial Memory: Remembering where, pp. 251–285. Lawrence Erlbaum Associates, Mahwah (2004)
11. Shelton, A.L., Gabrieli, J.D.: Neural correlates of encoding space from route and survey perspectives. Journal of Neuroscience 22, 2711–2717 (2002)
12. Gramann, K., Müller, H.J., Eick, E.M., Schönebeck, B.: Evidence of separable spatial representations in a virtual navigation task. Journal of Experimental Psychology - Human Perception and Performance 31, 1199–1223 (2005)
13. Gramann, K., Müller, H.J., Schönebeck, B., Debus, G.: The neural basis of ego- and allocentric reference frames in spatial navigation: Evidence from spatio-temporal coupled current density reconstruction. Brain Research 1118, 116–129 (2006)
14. Bohbot, V.D., Lerch, J., Thorndycraft, B., Iaria, G., Zijdenbos, A.P.: Gray matter differences correlate with spontaneous strategies in a human virtual navigation task. Journal of Neuroscience 27, 10078–10083 (2007)
15. Gramann, K., El Sharkawy, J., Deubel, H.: Eye-movements during navigation in a virtual tunnel. International Journal of Neuroscience 119, 1755–1778 (2009)
16. Galati, G., Lobel, E., Vallar, G., Berthoz, A., Pizzamiglio, L., Le Bihan, D.: The neural basis of egocentric and allocentric coding of space in humans: A functional magnetic resonance study. Experimental Brain Research 133, 156–164 (2000)
17. Merriam, E.P., Genovese, C.R., Colby, C.L.: Spatial updating in human parietal cortex. Neuron 39, 361–373 (2003)
18. Cavanna, A.E., Trimble, M.R.: The precuneus: A review of its functional anatomy and behavioural correlates. Brain 129, 564–583 (2006)
19. Andersen, R.A., Snyder, L.H., Bradley, D.C., Xing, J.: Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annual Reviews in Neuroscience 20, 303–330 (1997)
20. Aguirre, G.K., D'Esposito, M.: Topographical disorientation: A synthesis and taxonomy. Brain 122, 1613–1628 (1999)
21. Seubert, J., Humphreys, G.W., Müller, H.J., Gramann, K.: Straight after the turn: The role of the parietal lobes in egocentric space processing. Neurocase 14, 204–219 (2008)

22. Vann, S.D., Aggleton, J.P.: Testing the importance of the retrosplenial guidance system: Effects of different sized retrosplenial cortex lesions on heading direction and spatial working memory. Behavioural Brain Research 155, 97–108 (2004)
23. Burgess, N., Maguire, E.A., Spiers, H.J., O'Keefe, J.: A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. Neuroimage 14, 439–453 (2001)
24. Ino, T., Inoue, Y., Kage, M., Hirose, S., Kimura, T., Fukuyama, H.: Mental navigation in humans is processed in the anterior bank of the parieto-occipital sulcus. Neuroscience Letters 322, 182–186 (2002)
25. Redish, A.D.: Beyond the cognitive map: From place cells to episodic memory. The MIT Press, Cambridge (1999)
26. McNaughton, B.L., Battaglia, F.P., Jensen, O., Moser, E.I., Moser, M.B.: Path integration and the neural basis of the 'cognitive map'. Nature Reviews Neuroscience 7, 663–678 (2006)
27. Hafting, T., Fyhn, M., Molden, S., Moser, M.B., Moser, E.I.: Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806 (2005)
28. Maguire, E.A., Burgess, N., Donnett, J.G., Frackowiak, R.S., Frith, C.D., O'Keefe, J.: Knowing where and getting there: A human navigation network. Science 280, 921–924 (1998)
29. Jensen, O., Tesche, C.D.: Frontal theta activity in humans increases with memory load in a working memory task. European Journal of Neuroscience 15, 1395–1399 (2002)
30. Bischof, W.F., Boulanger, P.: Spatial navigation in virtual reality environments: An EEG analysis. CyberPsychology and Behavior 6, 487–495 (2003)
31. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. Brain Research Reviews 29, 169–195 (1999)
32. Lehmann, D., König, T.: Spatio-temporal dynamics of alpha brain electric fields, and cognitive modes. International Journal of Psychophysiology 26, 99–112 (1997)
33. Pfurtscheller, G., Lopez da Silva, F.H.: Event-related EEG/MEG synchronization and desynchronization: Basic principles. Clinical Neurophysiology 110, 1842–1857 (1999)
34. Onton, J., Makeig, S.: Information-based modeling of event-related brain dynamics. In: Neuper, C., Klimesch, E. (eds.) Progress in Brain Research, vol. 159, pp. 99–120 (2006)
35. Onton, J., Westerfield, M., Townsend, J., Makeig, S.: Imaging human EEG dynamics using independent component analysis. Neuroscience and Biobehavioral Reviews 30, 808–822 (2006)
36. Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.J.: Independent component analysis of electroencephalographic data. Advances in Neural Information Processing Systems 8, 145–151 (1996)
37. Oostenveld, R., Praamstra, P.: The five percent electrode system for high-resolution EEG and ERP measurements. Clinical Neurophysiology 112, 713–719 (2001)
38. Delorme, A., Makeig, S.: EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134, 9–21 (2004)
39. Lee, T.W., Girolami, M., Sejnowski, T.J.: Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Computation 11, 417–441 (1999)
40. Oostenveld, R., Oostendorp, T.F.: Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. Human Brain Mapping 17, 179–192 (2002)

41. Makeig, S.: Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalography and Clinical Neurophysiology 86, 283–293 (1993)
42. Makeig, S., Debener, S., Onton, J., Delorme, A.: Mining event-related brain dynamics. Trends in Cognitive Sciences 8, 204–210 (2004)
43. Gramann, K., Onton, J., Riccobon, D., Müller, H.J., Bardins, S., Makeig, S.: Human brain dynamics accompanying the use of egocentric and allocentric reference frames during spatial navigation. Journal of Cognitive Neuroscience (in press)
44. Gevins, A., Smith, M.E., McEvoy, L., Yu, D.: High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. Cerebral Cortex 7, 374–385 (1997)
45. Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T.: Automated Talairach atlas labels for functional brain mapping. Human Brain Mapping 10, 120–131 (2000)
46. Brodmann, K.: Vergleichende Lokalisationslehre der Grosshirnrinde: In ihren Principien dargestellt auf Grund des Zellenbaues. Johann Ambrosius Barth Verlag, Leipzig (1925)
47. Harris, I.M., Egan, G.F., Sonkkila, C., Tochon-Danguy, H.J., Paxinos, G., Watson, J.D.G.: Selective right parietal lobe activation during mental rotation - A parametric PET study. Brain 123, 65–73 (2000)
48. Morrone, M.C., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., Burr, D.C.: A cortical area that responds specifically to optic flow, revealed by fMRI. Nature Neuroscience 3, 1322–1328 (2000)
49. Calton, J.L., Taube, J.S.: Where am I and how will I get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. Neurobiology of Learning and Memory 91, 186–196 (2009)
50. Field, D.T., Wilkie, R.M., Wann, J.P.: Neural systems in the visual control of steering. Journal of Neuroscience 27, 8002–8010 (2007)
51. Andersen, R.A.: Encoding of intention and spatial location in the posterior parietal cortex. Cerebral Cortex 5, 457–469 (1995)
52. Klimesch, W., Doppelmayr, M., Schwaiger, J., Auinger, P., Winkler, T.: 'Paradoxical' alpha synchronization in a memory task. Cognitive Brain Research 7, 493–501 (1999)
53. Kahana, M.J., Sekuler, R., Caplan, J.B., Kirschen, M., Madsen, J.R.: Human theta oscillations exhibit task dependence during virtual maze navigation. Nature 399, 781–784 (1999)
54. Caplan, J.B., Madsen, J.R., Schulze-Bonhage, A., Aschenbrenner-Scheibe, R., Newman, E.L., Kahana, M.J.: Human theta oscillations related to sensorimotor integration and spatial learning. Journal of Neuroscience 23, 4726–4736 (2003)
55. Makeig, S., Gramann, K., Jung, T.P., Sejnowski, T., Poizner, H.: Linking brain, mind and behavior. International Journal of Psychophysiology 73, 95–100 (2009)

Putting Egocentric and Allocentric into Perspective

Tobias Meilinger[1] and Gottfried Vosgerau[2]

[1] Max-Planck-Institute for Biological Cybernetics
Spemannstr. 44, 72076 Tübingen, Germany
tobias.meilinger@tuebingen.mpg.de
[2] Institut für Philosophie, Heinrich-Heine-Universität Düsseldorf,
Universitätsstr. 1, 40225 Düsseldorf, Germany
gottfried.vosgerau@uni-duesseldorf.de

Abstract. In the last decade many studies examined egocentric and allocentric spatial relations. For various tasks, navigators profit from both kinds of relations. However, their interrelation seems to be underspecified. We present four elementary representations of allocentric and egocentric relations (sensorimotor contingencies, egocentric coordinate systems, allocentric coordinate systems, and perspective-free representations) and discuss them with respect to their encoding and retrieval. Elementary representations are problematic for capturing large spaces and situations which encompass both allocentric and egocentric relations at the same time. Complex spatial representations provide a solution to this problem. They combine elementary coordinate representations either by pair-wise connections or by hierarchical embedding. We discuss complex spatial representations with respect to computational requirements and their plausibility regarding behavioral and neural findings. This work is meant to clarify concepts of egocentric and allocentric, to show their limitations, benefits and empirical plausibility and to point out new directions for future research.

Keywords: spatial memory, egocentric, allocentric, sensorimotor contingencies, coordinate system, viewpoint-dependent, perspective-free, parietal cortex, hippocampus.

1 Introduction

The Spatial Cognition series started with a seminal paper by Roberta Klatzky ([20]) in which she provided a precise definition for egocentric (i.e., self-to-object) and allocentric (object-to-object) relations, namely self- or object-centered coordinate systems which define relative directions, distances, and bearings. This work has triggered much research examining ego- and allocentricity (e.g., [4], [47], [51]) and we have learned a great deal about these issues. In the following we will firstly summarize some of this work, secondly, compare Klatzky's definition with alternative conceptions of egocentric and allocentric representations with respect to encoding and retrieval, thirdly, discuss limitations of these conceptions, and finally, propose how to combine the existing conceptions in order to solve these limitations.

Humans are able to represent locations egocentrically as well as allocentrically. Depending on the circumstances they seem to prefer the exploitation of one kind of

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 207–221, 2010.

relation. Especially three spatial tasks were studied intensely for which we want to present a short overview: scene recognition, reorientation, as well as updating and disorientation.

When *recognizing scenes*, egocentric relations do play an important role. In most of the cases, scenes are recognized better based on an experienced view than on novel views ([9], [25], [37]). However, presenting a novel view interpolating two familiar views from close-by viewpoints can yield even better performance than presenting one of the familiar views ([48]). This suggests a combination of two egocentric representations. However, allocentric relations are also useful in scene recognition. Recognition is influenced by background objects ([5], [28]) or the intrinsic orientation of a spatial array ([27]).

The case is similar for *reorienting* (i.e., self-localizing one's position within a familiar environment after being disoriented). The geometric form of a room plays an important role for this task. This effect was mainly considered to arise from allocentric representations of the room ([4], [52]). However, simulations have shown that many of the experimental results can be explained by view matching as well which encompasses egocentric relations ([40]). Nevertheless, starting from five years of age, children begin reorienting by the (allocentric) structure of an object and not just by matching views ([29]).

A typical example of a task where egocentric relations are used is *updating*. When moving around, egocentric locations in the environment have to be updated. Although updating processes are always possible sources of errors, they are sufficiently accurate to detect changes to an object array equally well from a familiar viewpoint as from an updated novel viewpoint ([37]). However, navigators also encode allocentric relations. Since allocentric relations are independent from the own position, their representation does not need to be updated during own movement and is even preserved after disorientation. These allocentric relations encompass the shape of a surrounding room or a sufficiently familiar object array ([16], [47], [51]). Regular arrays seem to be represented more often allocentrically, whereas irregular layouts more likely egocentrically ([53]). When becoming more familiar with an array, a switch between precise egocentric to a more imprecise allocentric representation seems to take place ([47]).

The insights gained about scene recognition, reorientation and updating clearly show the usefulness of a precise definition of allocentric and egocentric relations. This definition allowed, first, identifying the format of spatial representations in experiments which thus, second, revealed under which circumstances navigators profit from which relation. However, egocentric and allocentric can be conceptualized not only as coordinate systems and need not be conceived of as mutually excluding conceptions. In the next section we will introduce two alternative formats of representation in addition to coordinate representations and will discuss how egocentric and allocentric representations may be encoded and retrieved.

2 Elementary Spatial Representations

An elementary conception is a spatial representation which, first, encompasses only one kind of relation (either egocentric or allocentric), and second, expresses locations relative to one (or no) point of reference. Therefore, egocentric and allocentric coordinate systems as introduced above are kinds of elementary spatial representations. In addition, we will now introduce sensorimotor representations (which form a second kind of elementary egocentric representation) and perspective-free representations (which form a second kind of elementary allocentric representation). We will then discuss processes of constructing and using these elementary representations and point out their limitations.

2.1 Egocentric Representations

2.1.1 Sensorimotor Contingencies

Following O'Regan and Noë[1] ([32], [33]; see also [45]) representing an object in a certain distance and direction means knowing how the sensory input will change when performing actions. It is the contingencies between action and perception within which a location is represented. Navigators know how their visual input on the retina will change when they move around. This change depends on the distance and the direction to the object. For example, distant locations result in smaller changes on the retinal picture when stepping one meter to the side than proximal locations. Sensorimotor contingencies are representations in a perception-action format. Navigators use them to deal with their environment, but not necessarily to communicate or think about it. As our perceptions and actions are always relative to our self (or to body parts), sensorimotor contingencies are egocentric representations.

Campbell ([8]) argued that "true" egocentric representations (in a philosophical sense) are to be distinguished from merely body-centered representations, the latter being representations where in principle the body could be any body (it only "happens to be mine"). Spatial representations based on sensorimotor contingencies are truly egocentric in this sense, since they represent spatial locations by (possible) actions (e.g. how to reach them or how to move to see them; see also [44], [45]). Since they represent actions (and not just spatio-temporal trajectories, which would make them merely body-centered), they do not encompass an explicit representation within a coordinate system, opposed to the following conception.

2.1.2 Egocentric Coordinates

In this conception introduced by Klatzky ([20]), spatial locations are explicitly represented within a coordinate system centered on the navigator (Fig. 1 left side). For navigation purposes, centers of coordinate systems will typically be the torso. Locations in a plane are specified by two parameters such as angle and distance relative to the current body orientation. In such a way each location is represented by an individual body-centered vector. Egocentric bearing (i.e., the angle between self-orientation and the orientation of another object) requires an additional parameter in the 2D case. Relations such as distance, relative direction, or relative bearing between two non-self

---

[1] The sensorimotor account of vision is an account of (conscious) visual experience by far not limited to spatial representations. It was worked out in detail in Noë ([30]). Although the authors do not agree with the general claim of O'Regan and Noë (namely that every conscious experience can be explained in this way; see also [35], [46] for critique), the basic idea can be fruitfully applied to spatial representation.

Fig. 1. Visualizations of egocentric (left) and allocentric (right) coordinate representations. In an egocentric coordinate system locations are represented relative to the body-orientation of a navigator as indicated by arrows. An allocentric coordinate system indicated by arrows and a grid represents locations as coordinates in a system centered on entities other than a navigator, such as an object array and/or the surrounding room.

locations (i.e., allocentric relations) are not directly represented, but can be derived from this representation. Egocentric coordinates can be stored in long term memory. When representing egocentric coordinates in working memory they have to be constantly updated during movements ([37]).

2.2 Allocentric Representations

2.2.1 Allocentric Coordinates

According to Klatzky ([20]), an allocentric coordinate system is located and oriented on an object or a location other than the navigator (Fig.1 right side). Stationary objects do not change their coordinates or bearings when the navigator (i.e., the representing system) is moving. Distance, direction and bearing of an object relative to the origin of the coordinate system are directly represented. Relations between two objects or locations other than the origin have to be derived. However, this is often assumed as computationally easy. The origin of an allocentric coordinate system has also been proposed as arbitrary or virtual ([13]). Please note that a pair-wise relation between two objects or locations is equivalent with the origin of a coordinate system centered on one of the two objects.

2.2.2 Perspective-Free Representations

An alternative interpretation of allocentric is non-centered. However, a coordinate representation is always centered, as the coordinate system has to have a defined origin with an orientation – even if the importance of this origin is de-emphasized. Structural descriptions offer a way to describe spatial relations in a non-centered way. For example, one can list all pair-wise distances between locations. Also the following constraints provide a structural description: distance(A,B) = distance(B,C) = 5 meters; angle(ABC) = 90°. If not given directly in the description, all relations have to be derived from the representation. Structural descriptions often have multiple solutions ([13]). Please note that many descriptions are perspective dependent and thus better described by egocentric or allocentric coordinates (e.g., “the trumpet is left of the hammer, the hammer is left of the teapot”).

2.3 The Encoding and the Retrieval of Elementary Spatial Representations

In the Introduction we summarized evidence for egocentric and allocentric coordinate representations. These and other spatial representations have to be formed from perception and are used for action. In the following we will present theoretic considerations of how these processes can be explicated along with empirical evidence for the presented conceptualization.

2.3.1 Deriving Egocentric (Trunk-)Coordinates from Lower-Level Representations

Our perceptions are relative to our own position, or even – more detailed – relative to the locations of parts of our body (e.g., relative to the positions of our hand, our head or our retina). All these representations are often called egocentric.[2] An egocentric trunk-based representation is derived from peripheral reference frames. Imagine using your hand for searching a table with closed eyes until you touch the object. By knowing where your hand is relative to your trunk you can derive the egocentric (trunk) position of the touched object. The visual identification of object locations works similarly. The location of an object on the retinas (i.e., in retinotropic coordinates) is transformed into head coordinates and then to egocentric trunk coordinates. Neurophsiological studies show that egocentric coordinate systems exist in the posterior parietal cortex ([1], [12]). Behavioral experiments indicate that head and trunk-based reference frames can be stored in memory ([50]). From the theoretical side, Grush ([13]) showed how a coordinate structure can be derived from lower level coupled sensory and action channels by a process he called s-coordination. This provides a model of how coordinate systems could be derived from a perception-action format as found in sensorimotor contingencies.

2.3.2 Deriving Allocentric Coordinates from Lower-Level Representations

Allocentric coordinates have to be derived from egocentric ones, just as egocentric coordinates are derived from data in a perception-action format. Deriving allocentric coordinates directly from a perception-action format would at least implicitly encompass egocentric relations. Byrne, Baker, and Burgess ([6]) suggested that such a transformation starts with egocentric coordinates in posterior parietal cortex, involves retrosplenial cortex, head-direction cells in the Papez circuit ([41]), and finally result in allocentric coordinates in the hippocampus (see also [11]). In hippocampal place cells a navigator’s location is represented relative to the immediate surrounding area ([31]). The navigator seems to be the only “object” represented within such an allocentric representation.

---

[2] As these representations are merely body-centered they cannot be classified as true egocentric representations in Campbell’s sense (see 2.1.1). Please note that retinal, head or torso-based coordinate systems each refer to one reference point (i.e., the retina, or the head, etc.) and thus are elementary spatial representations.

From a cognitive point of view, several solutions to the problem of transforming egocentric into allocentric coordinates are possible. One possibility is to mentally shift the perspective from which an environment is encoded to a non-ego position. The non-ego position works as the origin of a coordinate system relative to which locations are encoded. As this position was not (physically) encountered, the resulting coordinate system is an allocentric one. Humans are shown to be capable of using such a non-ego position ([26]).

Another possibility is to derive the allocentric relations directly from the egocentric coordinates. When wanting to cross a street with heavy traffic, one can represent the car locations by egocentric vectors. The allocentric relation 'the red car is in front of the silver car' can be computed from the egocentric representation. This representation can be in a natural language format as the example sentence or in a format closer to perception. In this way coordinate systems relative to any structure (objects, rooms, etc.) in the egocentrically represented environment can be derived.

Alternatively to an allocentric coordinate system also a perspective-free representation might be formed from egocentric representations. This would be a structural description of the environment (see 2.2.2). A perspective-independent behavior could also be obtained when assuming that the coordinate origin – especially its orientation – does not exhibit large computational and thus behavioral consequences. However, most studies do find performance differences which can be explained by the orientation of allocentric coordinate systems respectively relations. This indicates that the orientation of an allocentric reference frame does play a role ([23], [49]). Consequently, perspective-free representations are allocentric representations which only exist in addition to allocentric coordinate representations whose orientation matters.

The last sections examined how allocentric representations and egocentric coordinate representations can be formed from lower level representations and discussed empirical evidence for it. The next section is concerned with how such representations are retrieved and used for action.

2.3.3 Retrieving Elementary Spatial Representations

If egocentric and allocentric representations are to guide action, they have to be transformed into a sensorimotor format. Inverting the construction process of such representations is a potential model for how this might work. The coordinate system representing spatial relations in memory has to be transformed into the "current" egocentric coordinate systems of the navigator and from there into a perception-action format to elicit behavior. The former transformation is not required when all relevant information was updated or an egocentric long-term memory representation from the same viewpoint can be accessed. A differently oriented egocentric or any allocentric memory requires a coordinate transformation into the current egocentric orientation. Egocentric long-term representations from cleverly chosen view-points can thus minimize the transformations required during retrieval. For example, decision points in route navigation might be represented in this way ([25]). Structural descriptions also have to be transformed e.g., by building a mental model of the description (cf., [17], [18]).

Indeed, the described coordinate transformations can be shown to occur in alignment experiments. An alignment between a navigator's current orientation in a physical or imagined environment and the orientation of the egocentric or allocentric memory representation of this environment yields better performance than if the current body orientation and the memory orientation are misaligned ([9], [26], [37]). For allocentric representations, orthogonal misalignments seem to be less detrimental than oblique misalignments ([19], [26]).

When elementary spatial representations are used for guiding behavior, it must be clear which representation is guiding behavior at a certain time. This is unproblematic for representations of different locations: Only the representation of the environment crucial for the task at hand is used and not a representation of the distal environment. If during encoding only one kind of representation (egocentric or allocentric) is formed due to the circumstances, this is not problematic either. Only the existing representation guides behavior. However, it has been shown that participants can encode egocentric and allocentric representations of one and the same environment and use them depending on the task requirements ([43], [53]). Task specifications must thus select the appropriate representation. Alternatively, egocentric and allocentric representations could be combined. This, however, results in a non-elementary representation and will thus be discussed in section 3.

As shown so far, there is evidence for elementary spatial representations; their construction and usage can be made plausible both theoretically and empirically. However, not all situations can be captured successfully by elementary spatial representations.

2.4 Limitations of Elementary Representations

Explaining behavior by elementary spatial representations alone poses some difficulties for representing larger spaces as well as for representing objects, snapshots and some scenes.

2.4.1 Representing Larger Spaces

Representing environmental spaces such as cities or buildings with elementary representations only poses some specific problems. As more and more locations are encountered, representing them within a single representation will require bigger and bigger representations. For example, if all locations within a city would have to be represented within a single egocentric or allocentric representation, representations with an enormous amount of information would have to be dealt with. From a computational point of view our mental resources seem too limited to represent a city within a single coordinate system which would have to be accessed as a whole. Thus, substructures have to be formed which are not elementary representations in the sense used here. Alternatively, we could form multiple representations of local surroundings (e.g., each street) and use these representations when dealing with this street. Then, however, these local representations are unconnected. Elementary representations do not provide means of how to represent relations between multiple representations. It seems impossible to find a route to a goal or to point there based on unconnected local representations. In principle, this problem already occurs for visible spaces. Representing the location of each cobblestone by a single vector or by an allocentric coordinate requires very large memory. It seems reasonable to cluster locations together e.g., to form a floor, wall or house. However, such clustering requires an extension of the so far described elementary representations.

2.4.2 Situations Which Require Egocentric and Allocentric Relations to Represent Them

Some examples are difficult to confirm to elementary representations, because they seem to rely on allocentric as well as egocentric relations at the same time. Pictorial snapshots are often considered typical examples for egocentric representations ([52]). However, probably only very few people would conceptualize a snapshot as vectors pointing to each pixel of the snapshot. The snapshot surely is represented from a certain perspective; however, the pixels are also represented relative to the picture frame, which is an allocentric relation. The same applies to the memory of a (paper) map. Relations between locations on this map are memorized (i.e., allocentric relations). However, the map is also encoded in the orientation it was perceived which is an egocentric relation ([42]). This egocentricity cannot be captured by just declaring the whole memory as allocentric, because the reference point (i.e., the origin of the allocentric coordinate system) is unclear. Is it the border of the snapshot or map or is it the observer? Both have to be taken into account which is not possible within a single elementary representation. Broadening the definition of egocentric to all representations taken from an experienced location is not a good solution either. Then egocentric representations would also encompass object-to-object relations and the egocentric-allocentric distinctions would be reduced to representing an environment from an experienced or not experienced viewpoint.

Also object parts are represented relative to an object based reference frame which is an allocentric relation. For example, we surely do not represent mere egocentric vectors to object parts and update them individually. We do not err about how object parts relate to each other, even after disorientation. Despite this, objects seem to be encoded from experienced views ([3]). To explain this, an additional reference has to be taken into account – namely the observer. This is an egocentric relation. Consequently, also the representation of objects requires egocentric and allocentric relations together. In terms of object processing, both view-dependent and structural elements have to be considered ([14]).

A last example is the verbal statement "the target is left of the tree". This statement involves three locations: of the speaker, the target and the tree. Similarly to the examples above this situation cannot be sufficiently captured by egocentric vectors only. The relation between target and tree is one between two non-egocentric locations, therefore, an allocentric relation. On the other hand, the situation can neither be conceptualized in a mere allocentric way. As the tree does not offer any directional cues by itself, the direction of the target relative to the tree (i.e., left of) is derived from the perspective of the speaker which is an egocentric relation. Again, egocentric and allocentric elements seem to be combined.

We propose that solutions to the mentioned problems require combining elementary spatial representations into complex representations. In the following chapter we will propose three ways of how such complex spatial representations can be conceptualized.

3 Complex Spatial Representations

All four elementary representations can be extended to represent further locations by simply adding more locations within the representation. As indicated in the last section this does not seem sufficient to explain all spatial behavior. The allocentric and the egocentric coordinate representations (2.1.2 and 2.2.1) can also be combined with each other in a hierarchical way or by pair-wise relations forming a complex representation.

3.1 Pair-Wise Relations

A pair-wise relation between two coordinate systems specifies the direction and distance necessary to get from coordinate system 1 to coordinate system 2 as well as the angle between the orientations of the two systems. In mathematical terms this is a matrix multiplication. In psychological terms this corresponds to a perspective shift. Due to the equivalence of a pair-wise relation and a coordinate representation, a pair-wise relation can be instantiated by considering the origin of coordinate system 2 as a position (location plus orientation) within coordinate system 1. As this can work the opposite direction at the same time, both coordinate systems are not related in a hierarchical fashion, but are on the same level. Pair-wise relations could occur between egocentric, allocentric and combined ego- and allocentric coordinate systems. Pair-wise relations can avoid the problem of representing many locations, as only one coordinate system or a subset can be selected and used for orientation in working memory. This selection is possible as complex spatial representations are compositional (i.e., they are combined out of elementary spatial representations). Still, the granularity of the spatial representation does matter. Pair-wise relations between representations of vista spaces such as streets or rooms do seem plausible ([24]). However, connecting all objects within a city or a building by pair-wise relations requires many relations to be encoded. This does not seem very practical from a computational point of view. A limitation to represent only relations to neighbors can solve this problem. Hierarchical representations provide an alternative solution.

3.2 Hierarchical Relations

3.2.1 Allocentric Hierarchy

Hierarchical conceptions of spatial memory are popular ([15], [23], [39]). In this specific conception lower level coordinate systems (ego- or allocentric) are integrated within a top-level allocentric coordinate system (Fig. 2 left side). The area of the lower level coordinate system is thus part of the area represented by the higher level coordinate system. This is not necessarily the case for pair-wise relations. A top-level coordinate system could correspond, for example, to a city whereas lower level coordinate systems could represent single streets within this city in more detail. Or the top-level coordinate system corresponds to a room and the lower level one to objects within this room. The lower level coordinate systems are not necessarily allocentric, but could be egocentric as well.

A specific example for such an allocentric-egocentric combination is the position of a navigator (i.e., an egocentric coordinate system) within an allocentric map ([6], [36]). In such a representation the position of the navigator relative to an environment is constantly updated during movement. A potential problem of such an updating is the choice of the relevant hierarchy level relative to which the egocentric coordinate systems should be updated. How does one decide if it is relative to the city, the district or the street level, or relative to all levels at the same time? In addition, what happens when moving from one street to the next one (which is covered in detail by another representation)? This would have to be specified in more detail.

Survey and route knowledge can also be conceptualized within one allocentric hierarchy (see also [34]). Pair-wise relations between lower level coordinate systems represent route knowledge (i.e., knowing in which direction the next reference frame is located). These lower level coordinate systems are embedded within a higher-level allocentric coordinate system within which the relative spatial locations between distant lower-level coordinate systems are specified.

Elements in a hierarchy might also be linked to non-spatial information (see also [22]). For example, labels such as "city hall", "downtown", or "Tübingen" can be attached to one or multiple coordinate systems ([24]). These labels can relate to general background knowledge about the world.

Hierarchical relations between coordinate systems easily avoid the problem of large spaces. Only the hierarchical level necessary for the task at hand has to be accessed thus avoiding too large data sets to be processed.

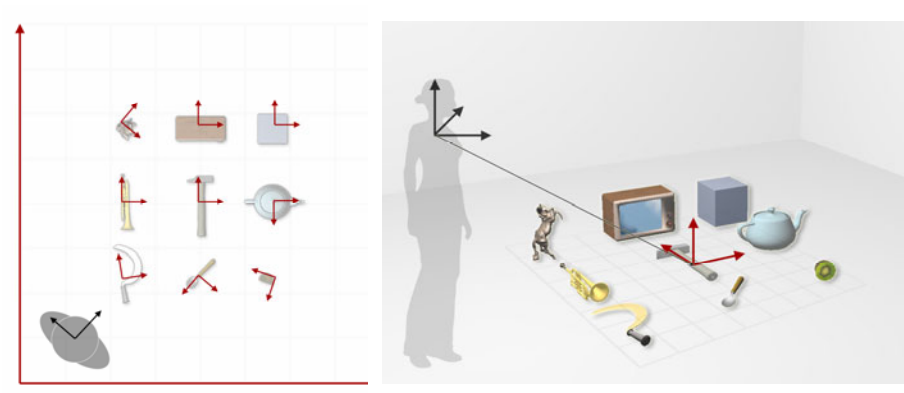


Fig. 2. Visualizations of hierarchical spatial representations. In an allocentric hierarchy (left side) the top level allocentric reference frame (long arrows and grid) defines lower level allocentric (here for each object) or egocentric (the navigator) reference frames. In an egocentric hierarchy (right side) lower level reference frames (e.g., for a room, an object array as shown here, or for individual objects) are subsumed under a top-level egocentric reference frame.

3.2.2 Egocentric Hierarchy

In this conception, lower level egocentric or allocentric coordinate systems are subsumed under a top-level egocentric reference frame (Fig. 2 right side). For example, an allocentric reference frame of an object is subsumed under an egocentric one. The object is represented from a certain perspective; however, also the allocentric relations between object parts are represented. Such a representation captures view-dependent and structural elements which both have been shown to contribute to object recognition ([14]). Similarly, the double nature of memorized views or maps can be captured. Locations on the snapshot or map are represented allocentrically relative to the map or snapshot frame. This frame is part of an egocentric representation capturing the perspective element. In the tree example, the allocentric coordinate system of the tree inherits its orientation from the higher level egocentric one thus specifying the relation left of the tree. Please note that also allocentric hierarchies are technically capable of solving this problem. The object, view, map or target-tree line defines the top-level allocentric coordinate system and the observer is subsumed as a lower-level egocentric system. Intuitively, this might not seem to capture the situations adequately, but from a formal point of view it is sufficient.

The problem of representing large spaces is solved via the hierarchical structure of the representation just as in the case of allocentric hierarchy. The character of the top-level reference frame is irrelevant for this; it is the hierarchical structure which is crucial. However, the hierarchical embedding under an egocentric top-level reference frame has the advantage that "large-scale" behavior (navigation) can be triggered directly by the egocentric top-level representation without costly transformations. If needed, however, detail information can be retrieved from embedded allocentric (or egocentric) representations and used – after transformation – as an additional source for guidance.

Updating an egocentric hierarchy during movements does not cause the same reference problems as in the case of allocentric hierarchy. A limited working memory buffer keeping a few objects could be updated. New visual input would overwrite existing objects (i.e., updated coordinate systems) unless rehearsal processes such as attention shift protect them.

Also route and survey knowledge can be conceptualized within an egocentric hierarchy. Route knowledge may be represented by pair-wise related coordinate systems just as in an allocentric hierarchy. This long-term knowledge might then be accessed from a currently active top-level egocentric reference frame and navigational instruction be derived. Rather than representing survey relations explicitly within a higher-level allocentric coordinate system, they can be derived online from the long-term representation. This can be achieved by integrating the pair-wise relations between coordinate systems along a route to a target within the currently active egocentric top-level coordinate system (for details see [24]).

The conscious perception of our surrounding world can also be conceptualized within an ego-allocentric hierarchy: an egocentric stage containing allocentric elements such as the form of the room or objects within ([45]). Attention might be focused on elements in this hierarchy either top-level or specific lower level coordinate frames ([21]) emphasizing egocentric or allocentric relations. All examples of egocentric hierarchies mentioned so far spanning from objects, snapshots, and maps, via object constellations and routes to survey relations can be consciously experienced or imagined. It can thus be conjectured that egocentric hierarchies might best capture our perspectival conscious experience of space.

Plausible brain areas corresponding to an egocentric hierarchy encompass posterior parietal cortex and the parahippocampal place area. As mentioned earlier, egocentric short term representations of a real or imagined current surrounding can be found in posterior parietal cortex ([2], [7], [12]). However, also neurons responding to allocentric relations have been observed ([38]). Although their interrelation was not specified, it is plausible to assume that they are linked in the sense of an egocentric hierarchy. Longer-term memory representations of an egocentric hierarchy for scenes seem to be found in the parahippocampal place area ([10], [11]). This area is more active when showing pictures of rooms or room parts (i.e., walls, floor, etc) in their correct arrangement than when scrambling the order of the room parts and thus destroying their allocentric relations ([10]). In addition, these representations are viewpoint dependent and, thus, sensitive to egocentric and allocentric aspects ([11]). Here again an egocentric hierarchy is plausible.

Complex spatial representations are capable of representing situations difficult to explain by elementary spatial representations only. Especially hierarchical solutions seems plausible from an empirical, a phenomenal as well as from a computational perspective.

4 Conclusions

Starting with the seminal work of Klatzky ([20]) many studies examined the application of egocentric and allocentric reference frames. The clear definition of these terms allowed for experiments distinguishing between these terms. In the present paper we compared the original definition of Klatzky with two alternative conceptions (i.e., sensorimotor contingencies and perspective-free representations) and discussed their encoding and retrieval for navigation. Although, the spatial cognition community has gained a lot of insights into when and how humans apply egocentric and allocentric references, not all phenomena can be captured with elementary conceptions of egocentric and allocentric representations. Especially two situations seem problematic: the representation of a large number of locations and representing objects and situations which seem to encompass egocentric and allocentric relations at the same time. In order to solve these problems, we propose complex spatial representations which are constructed from egocentric and allocentric coordinate systems. They can take multiple reference points into account and their compositionality is suited for a short-term memory of limited capacity. From our point of view, especially hierarchical conceptions and here especially hierarchies with a top-level egocentric coordinate system seem promising for representing also complex spatial situations. The discussion in the last years was shaped by a clear-cut distinction between egocentric and allocentric reference frames as separate systems of representation. We think that future research should not emphasize the separation of these representations, but rather their interaction.

Acknowledements. This research was supported by the DFG grant “The functional, computational and neural basis of human survey knowledge – comparing mental maps and mental graphs”. We would like to thank Heinrich Bülthoff for supporting this work, Stephan Streuber for help with the artwork, Betty Mohler for proofreading, and three anonymous reviewers for their helpful comments.

References

1. Andersen, R.A., Essick, G.K., Siegel, R.M.: Encoding of spatial location by posterior parietal neurons. Science 230, 456–458 (1985)
2. Andersen, R.A., Snyder, L.H., Bradley, D.C., Xing, J.: Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annual Reviews in Neuroscience 20, 303–330 (1997)

3. Bülthoff, H.H., Edelman, S.: Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proceedings of the National Acadamy of Sciences of the United States of America 89, 60–64 (1992)
4. Burgess, N.: Spatial memory: how egocentric and allocentric combine. Trends in Cognitive Sciences 10, 551–557 (2006)
5. Burgess, N., Spiers, H.J., Paleologou, E.: Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences on spatial memory. Cognition 94, 149–166 (2004)
6. Byrne, P., Becker, S., Burgess, N.: Remembering the past and imagining the future: a neural model of spatial memory and imagination. Psychological Review 114, 340–375 (2007)
7. Calton, J.L., Taube, J.S.: Where am I and how will I get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. Neurobiology of Learning and Memory 91, 186–196 (2009)
8. Campbell, J.: The role of physical objects in spatial thinking. In: Eilan, N., McCarthy, R., Brewer, B. (eds.) Spatial Representations. Problems in Philosophy and Psychology, pp. 65–95. Blackwell, Oxford (1993)
9. Diwadkar, V.A., McNamara, T.P.: Viewpoint dependence in scene recognition. Psychological Science 8, 302–307 (1997)
10. Epstein, R., Kanwisher, N.: A cortical representation of the local visual environment. Nature 392, 598–601 (1999)
11. Epstein, R.A.: The cortical basis of visual scene processing. Visual Cognition 12, 954–978 (2005)
12. Graziano, M.S.A., Cooke, D.F., Taylor, C.S.R.: Coding the location of arm by sight. Science 290, 1782–1786 (2000)
13. Grush, R.: Self, World and Space: The Meaning and Mechanisms of Ego- and Allocentric Spatial Representations. Brain and Mind 1, 59–92 (2000)
14. Hayward, W.G.: After the viewpoint debate: where next in object recognition? Trends in Cognitive Sciences 7, 425–427 (2003)
15. Hirtle, S.C., Jonides, J.: Evidence of hierarchies in cognitive maps. Memory & Cognition 13, 208–217 (1985)
16. Holmes, M.C., Sholl, M.J.: Allocentric coding of object-to-object relations in overlearned and novel environments. Journal of Experimental Psychology: Learning, Memory and Cognition 31, 1069–1078 (2005)
17. Jahn, G., Knauff, M., Johnson-Laird, P.N.: Preferred mental models in reasoning about spatial relations. Memory & Cognition 35, 2075–2087 (2007)
18. Johnson-Laird, P.N.: Mental models. Cambridge University Press, Cambridge (1983)
19. Kelly, J.W., McNamara, T.P.: Spatial memories of virtual environments: How egocentric experience, instrinsic structure, and extrinsic structure interact. Psychonomic Bulletin & Review 15, 322–327 (2008)
20. Klatzky, R.L.: Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In: Freksa, C., Habel, C., Wender, K.F. (eds.) Spatial Cognition 1998. LNCS (LNAI), vol. 1404, pp. 1–17. Springer, Heidelberg (1998)
21. Kuipers, B.: Drinking from the firehose of experience. Artificial Intelligence in Medicine 44, 155–170 (2008)
22. Mallot, H.A., Basten, K.: Embodied spatial cognition: Biological and artificial systems. Image and Vision Computing 27, 1658–1670 (2009)
23. McNamara, T.P., Slucenski, J., Rump, B.: Human Spatial Memory and Navigation. In: Roedinger III, H.L. (ed.) Cognitive Psychology of Memory. Learning and Memory: A comprehensive Reverence, vol. 2, 4 vols. (J. Byrne Editor). Elsevier, Oxford (2008)

24. Meilinger, T.: The network of reference frames theory: A synthesis of graphs and cognitive maps. In: Freksa, C., Newcombe, N.S., Gärdenfors, P., Wölfl, S. (eds.) Spatial Cognition VI. LNCS (LNAI), vol. 5248, pp. 344–360. Springer, Heidelberg (2008)
25. Meilinger, T., Franz, G., Bülthoff, H.H.: From Isovists via Mental Representations to Behaviour: First Steps Toward Closing the Causal Chain. Environment and Planning B (in press), doi:10.1068/b34048t
26. Mou, W., McNamara, T.P.: Intrinsic frames of reference in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 28, 162–170 (2002)
27. Mou, W., Fan, Y., McNamara, T.P., Owen, C.B.: Intrinsic frames of reference and egocentric viewpoints in scene recognition. Cognition 106, 750–769 (2008)
28. Mou, W., Xiao, C., McNamara, T.P.: Reference directions and reference objects in spatial memory of a briefly viewed layout. Cognition 108, 136–154 (2008)
29. Nardini, M., Thomas, R.L., Knowland, V.C.P., Braddick, O.J., Atkinson, J.: A viewpoint-independent process for spatial reorientation. Cognition 112, 241–248 (2009)
30. Noë, A.: Action in Perception. MIT Press, Cambridge (2005)
31. O'Keefe, J., Nadel, L.: The hippocampus as a cognitive map. Clarendon Press, Oxford (1978)
32. O'Regan, J.K., Noë, A.: A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences 22, 939–973 (2001a)
33. O'Regan, J.K., Noë, A.: What It is Like to See: A Sensorimotor Account of Vision and Visual Consciousness. Synthese 192, 79–103 (2001b)
34. Poucet, B.: Spatial cognitve maps in animals: new hypotheses on the structure and neural mechanisms. Psychological Review 100, 163–182 (1993)
35. Schlicht, T., Pompe, U.: Rezension von Alva Noë: Action in Perception. Zeitschrift für philosophische Forschung 61, 250–254 (2007)
36. Sholl, M.J., Nolin, T.L.: Orientation specificity in representations of place. Journal of Experimental Psychology: Learning, Memory, and Cognition 23, 1494–1507 (1997)
37. Simons, D.J., Wang, R.F.: Perceiving real-world viewpoint changes. Psychological Science 9, 315–320 (1998)
38. Snyder, L.H., Grieve, K.L., Brotchie, P., Anderson, R.A.: Separate body- and world-referenced representations of visual space in parietal cortex. Nature 394, 887–891 (1998)
39. Stevens, A., Coupe, P.: Distortions in judged spatial relations. Cognitive Psychology 10, 422–437 (1978)
40. Stürzl, W., Cheung, A., Cheng, K., Zeil, J.: The information content of panoramic images I: The rotational errors and the similarity of views in rectangular experimental arenas. Journal of Experimental Psychology: Animal Behavior Processes 34, 1–14 (2008)
41. Taube, J.S.: The head direction signal: origins and sensory-motor integration. Annual Review of Neuroscience 30, 181–207 (2007)
42. Tlauka, M., Nairn, M.J.: Encoding of multiple map orientations. Spatial Cognition and Computation 4, 359–372 (2004)
43. Valiquette, C., McNamara, T.P.: Different mental representations for place recognition and goal localization. Psychonomic Bulletin & Review 14, 676–680 (2007)
44. Vosgerau, G.: Conceptuality in Spatial Representation. Philosophical Psychology 20, 349–365 (2007)
45. Vosgerau, G.: Mental Representation and Self-Consciousness. From Basic Self-Representation to Self-Related Cognition. Mentis, Paderborn (2009)
46. Vosgerau, G., Schlicht, T., Newen, A.: Orthogonality of Phenomenality and Content. American Philosophical Quarterly 45, 309–328 (2008)

47. Waller, D., Hodgson, E.: Transient and enduring spatial representations under disorientation and self-rotation. Journal of Experimental Psychology: Learning, Memory, and Cognition 32, 867–882 (2006)
48. Waller, D., Friedman, A., Hodgson, E., Greenauer, N.: Learning scenes from multiple views: Novel views can be recognized more efficiently than learned views. Memory & Cognition 37, 90–99 (2009)
49. Waller, D., Montello, D.R., Richardson, A.E., Hegarty, M.: Orientation specificity and spatial updating of memories for layouts. Journal of Experimental Psychology: Learning, Memory, and Cognition 28, 1051–1063 (2002)
50. Waller, D., Lippa, Y., Richardson, A.: Isolating observer-based reference directions in human spatial memory: Head, body and the self-to-array axis. Cognition 106, 157–183 (2008)
51. Wang, R.F., Spelke, E.S.: Updating egocentric representations in human navigation. Cognition 77, 215–250 (2000)
52. Wang, F.R., Spelke, E.S.: Human spatial representation: insights form animals. Trends in Cognitive Sciences 6, 376–382 (2002)
53. Xiao, C., Mou, W., McNamara, T.P.: Use of self-to-object and object-to-object spatial relations in locomotion. Journal of Experimental Psychology: Learning, Memory, and Cognition 35, 1137–1147 (2009)

Reference Frames Influence Spatial Memory Development within and Across Sensory Modalities

Jonathan W. Kelly[1], Marios N. Avraamides[2], and Timothy P. McNamara[3]

[1] Department of Psychology, Iowa State University
W112 Lagomarcino Hall, Ames, IA 50010
[2] Department of Psychology, University of Cyprus
P.O. Box 20537, 1678 Nicosia, Cyprus
[3] Department of Psychology, Vanderbilt University
111 21st Ave. South, Nashville, TN 37203
jonkelly@iastate.edu, mariosav@ucy.ac.cy,
t.mcnamara@vanderbilt.edu

Abstract. Research on spatial memory indicates that locations are remembered relative to reference frames, which define a spatial reference system. Reference frames are thought to be selected on the basis of environment-based and experience-based cues present during learning. Results from new experiments indicate that reference frames provide scaffolding during the development of spatial memories: the reference frame used to organize locations studied from one perspective was also used to organize new locations studied from another perspective. Further results indicate that the role of reference frames during spatial memory development can cross sensory modalities. Reference frames that organized memories of a visually-experienced environment also organized memories of haptically-experienced locations studied within the same environment. These findings indicate a role for reference frames during spatial memory development, and demonstrate that reference frames influence cross-modal spatial learning.

Keywords: Reference frames, Spatial memory development, Perspective taking, Multi-modal learning.

1 Introduction

Spatial memories are critical to everyday navigation. Simple tasks such as locating one's keys before leaving home and more complex tasks such as finding a new route to avoid traffic congestion all depend on the navigator's ability to recall spatial memories of those environments. Mounting evidence indicates that spatial memories are orientation dependent, with privileged access during recall to orientations aligned with the reference frame in which the memory was stored (see McNamara, 2003, for a review). Reference frames are thought to be selected on the basis of cues present during learning, including environment-defined cues such as city street grids and experience-defined cues such as studied perspectives. The relative roles of those cues in selecting reference frames are discussed in more detail in Section 2.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 222–233, 2010.

Much of the research investigating reference frames in spatial memory has been conducted using environments in which the to-be-learned spatial layout is visible in its entirety from all learning perspectives. In contrast, natural spatial learning often involves incremental exposure to different parts of a scene or an environment. For example, someone relocating to a new neighborhood might first learn the locations of other homes and businesses on the same block as their new residence. Over time they explore the neighborhood and experience different views of their own street as well as other streets that they had not previously seen. After a couple of weeks they will have constructed a spatial memory of the neighborhood without ever experiencing the entire environment from a single view. Spatial memories of larger environments such as parks and neighborhoods, which are commonly learned in this piecemeal fashion, have also been shown to be orientation-dependent (McNamara, Rump & Werner, 2003; Montello, 1991; Werner & Schmidt, 1999).

Recent work from our labs has investigated the role of reference frames during the acquisition and development of spatial memories when learning occurs incrementally across multiple views. This research focuses on whether parts of a spatial layout learned from different views are remembered using different local frames of reference or whether they are integrated into a single global reference frame. These experiments, described in Section 3, indicate that reference frames established from an initial learning view can be used to organize new locations learned from other views, such that all locations are remembered within the same global reference frame. Further experiments indicate that reference frames can cross sensory modalities, whereby a reference frame used to remember a set of visually-learned objects also influences the reference frame used to remember another set of objects experienced later through touch. These findings establish a role for reference frames during the development of spatial memories. The findings also indicate that spatial reference frames may be amodal, and not specific to any individual sensory modality.

2 Reference Frames in Spatial Memory

Spatial locations are necessarily relative (e.g., the sink is *left of* the refrigerator, Iowa is *west* of Illinois, etc.), and so memories for spatial layouts must be stored in the context of a spatial reference system. Research indicates that spatial memories are commonly organized around allocentric reference frames centered on the environment (Avraamides & Kelly, 2005, 2008; Hintzman, O'Dell and Arndt, 1981; Kelly, Avraamides & Loomis, 2007; Kelly & McNamara, 2008; McNamara, 2003; McNamara, Rump & Werner, 2003; Montello, 1991; Mou & McNamara, 2002; Shelton & McNamara, 2001; Valiquette, McNamara & Smith, 2003; Werner & Schmidt, 1999; but see Wang & Spelke, 2000, 2002; Waller, Lippa & Richardson, 2008; Waller, Montello, Richardson & Hegarty, 2002, for evidence of egocentric reference frames centered on the body, head, eyes, etc.).

Shelton and McNamara (2001) conducted a series of studies investigating spatial memory organization. In a paradigmatic experiment, participants viewed seven objects placed on a square mat surrounded by a rectangular room. The edges of the mat were parallel with the room walls. Participants studied the objects from two views separated by 135°, and all seven objects were visible from both views. One view was aligned and one was misaligned with the environmental axes defined by the mat and the walls. Half of the participants studied from the aligned view first and from the misaligned view second and half of the participants studied in the opposite order. After studying, participants were led to another room for the retrieval test. On each trial, participants were asked to imagine standing at one object, facing a second object, and to point to a third object from that imagined perspective. Participants imagined different perspectives spaced evenly (every 45°) around the layout. Regardless of the viewing order during learning, perspective-taking performance was best when imagining the aligned study perspective and performance on the misaligned study perspective was no better than on other perspectives that were never experienced. Those results contrast with a subsequent experiment in which the mat on the floor was misaligned with the room walls, thereby disrupting the global environmental axes. Participants again studied from two views separated by 135°: one view was aligned with the axes of the room and the other view was aligned with the axes of the mat. During subsequent perspective taking, performance when imagining the two experienced views was better than non-experienced views, and performance on the first learning view was better than the second view, regardless of its alignment with the environmental cues.

Based on those studies, Shelton and McNamara (2001) proposed that spatial relations are stored within spatial reference frames selected on the basis of cues available during learning. Salient environmental cues (such as the mat on the floor aligned with the room walls) result in reference frames aligned with those environmental structures, especially when learning occurs from a view aligned with the environmental axes. When consistent environmental cues are lacking (as when the mat on the floor was misaligned with the room walls) egocentric cues dominate and reference frames are selected from the initially experienced view. But in both cases, the reference frame is thought to be fixed with respect to the environment, and therefore allocentric in nature.

Mou and McNamara (2002) presented evidence that reference frames can be intrinsic to the remembered layout. In their experiments, participants studied an object layout with a bilateral symmetry axis. The bilateral symmetry axis defined a salient environmental cue. After learning, participants were better able to imagine perspectives aligned with the intrinsic axis of the layout. This was true even when participants studied from a view misaligned with the symmetry axis of the layout.

Based in part on the evidence reviewed above, McNamara and colleagues have proposed that locations are remembered in the context of a spatial reference frame, selected on the basis of cues available during learning. Those cues can be broadly categorized as egocentric (experience-defined) cues, extrinsic (environment-defined) cues, and intrinsic (layout-defined) cues.

3 Reference Frames in Spatial Memory Development

Most of the work described in Section 2 focused on spatial memories of environments in which the entire environment was visible from all studied perspectives. In contrast, most real-world spatial learning occurs piecemeal, often due to the presence of occluding obstacles such as walls and tables in smaller environments and buildings, trees and hills in larger environments. The role of reference frames is unclear when environments are learned in this piecemeal manner. One possibility is that different parts of the environment, visible from different views, are encoded in separate local reference frames. Another possibility is that those different parts of the same environment are encoded in the same global reference frame.

One way in which this question has been addressed is by asking participants to learn two separate routes or layouts and later showing participants how the two spaces relate to one another. Subsequent judgments of between-layout and within-layout pairs of locations have been used as an indication of whether the two routes were represented in separate reference frames or in a common reference frame. For example, Hanley and Levine (1983) had participants learn two separate layouts on a tabletop environment. Each layout consisted of three points (points A, B, and C in one layout and points 1, 2, and 3 in the other layout) connected by two lines, forming V-shape patterns with different orientations. After learning both layouts, participants were shown that the two layouts actually contacted one another, such that point C from one of the V-shapes was spatially coincident with point 2 from the other V-shape. After studying the relationship between the two layouts, participants made judgments of distance and direction between pairs of points within the same layout or across the two layouts. The latter judgments required integration across the two layouts. If the layouts were stored in separate reference frames established during learning, then judgments should have been better for within- compared to between-layout judgments, since between-layout judgments would require mental rotation to bring the two layouts into alignment (e.g., Shepard & Metzler, 1971). In contrast, if the two layouts were integrated into a single reference frame then judgments should have been similar for within- and between-layout judgments. The data indicated that participants represented the layouts in separate reference frames, with superior performance on within-layout judgments. Superior within-layout performance has since been shown in other experiments varying in the learning conditions and environmental sizes (Golledge, Ruggles, Pellegrino & Gale, 1993; Ishikawa & Montello, 2006; Montello & Pick, 1993).

However, comparison of within- and between-layout judgments is complicated by the finding that spatial judgments can be influenced by both the temporal and spatial separation between objects during learning (McNamara, Halpin & Hardy, 1992), and this effect could be independent of the actual reference frame used to represent those locations. Further complicating matters are the discrepant results from experiments showing similar performance for within-layout and between-layout judgments, suggesting that layouts learned separately are, under some circumstances, integrated into a single reference frame (Holding & Holding, 1988; Maguire, Burke, Phillips & Staunton, 1996; Moar & Carleton, 1982). These discrepant findings could potentially be accounted for by differences in experimental design, including differences in temporal and spatial separation of layouts and differences in the cues known to influence reference frame selection (reviewed above in Section 2). Even more challenging to explain is the finding that switching between two imagined perspectives is faster when the two perspectives are from different environments compared to when they are from the same environment (Brockmole & Wang, 2002). However, this could be due to interference during the switching process (e.g., May, 1996, 2004) rather than reference frames per se.

In light of the methodological challenges associated with comparing within- and between layout judgments, we have recently developed a new method to address the influence of reference frames during the development of spatial memories. In one such experiment, participants learned two overlapping spatial layouts, shown in Figure 1. The layouts were surrounded by a circular room in order to limit the number of environmental cues that might influence reference frame selection. First, participants studied layout one (object names drawn with an oval border in Figure 1) in isolation from 0° or from 135°. Next, layout two (object names drawn without a border in Figure 1) was added to the environment. All participants studied layout two from the 135° view only. Layout one was present throughout learning, even when studying layout two. After learning, participants performed a perspective-taking task in which they imagined standing at one object, facing a second object, and then pointed to a third object from that imagined perspective. Triplets of objects comprising a single trial always came from the same layout (i.e., all three objects were drawn from layout one or all three objects were drawn from layout two). Both layouts were tested, but a single trial never mixed objects between layouts. Of critical importance was the comparison of perspective-taking performance for layout two across the layout one learning conditions. Any changes in the pattern of layout two perspective-taking performance would indicate that participants' memories of layout two were influenced by their experiences when learning layout one.

As expected, perspective-taking performance for trials testing layout one (Figure 2, left panel) depended on the layout one learning condition. Participants who first studied layout one from 0° were better able to recall layout one from perspectives of 0°, 90°, 180° and 270°, compared to 45°, 135°, 225° and 315°, producing the sawtooth

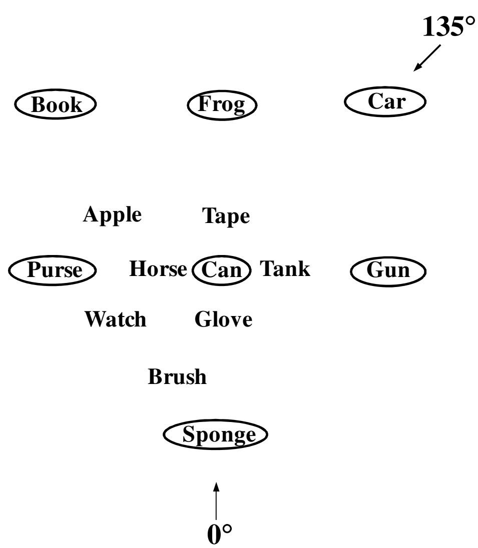


Fig. 1. Illustration of the two overlapping spatial layouts. Layout one (object names with oval borders) was studied from either the 0° view or the 135° view. Layout two (object names with no border) was then added to layout one. Layout two was always studied from the 135° view. The environment was surrounded by a circular room (not shown).

pattern seen in Figure 2. This finding of facilitated performance along orthogonal reference directions has been reported previously (e.g., Mou & McNamara, 2002). Participants who studied layout one from 135° were better able to retrieve layout one from 135° and 315° (the perspective diametrically opposite the studied perspective). The layout one results were expected based on previous research showing the influence of the learning perspective (e.g., Shelton & McNamara, 2001), and they serve as a benchmark for interpreting the results from layout two.

Layout two perspective-taking performance (Figure 2, right panel) also depended on the layout one learning condition. Participants who studied layout one from 0° also showed sawtooth pattern when recalling layout two, with facilitated performance from 0°, 90°, 180° and 270°, and also the 135° perspective from which layout two was studied. In contrast to the sawtooth pattern found when layout one was studied from 0°, participants who studied layout one from 135° showed superior performance when imagining layout two from 135°, with some facilitation also at 315°.

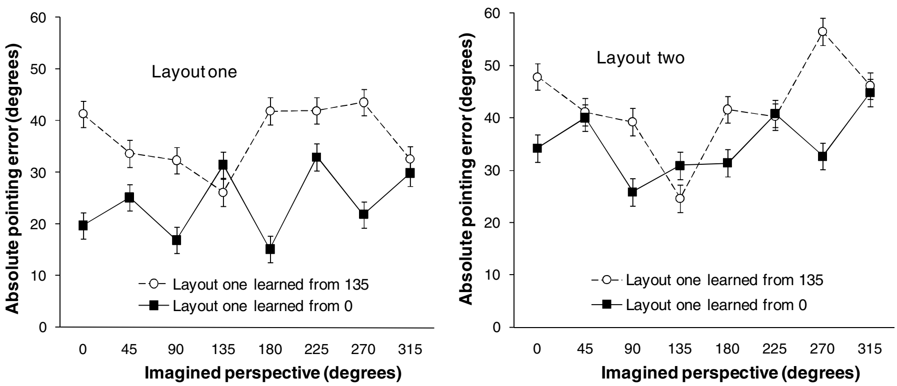


Fig. 2. Absolute pointing error during an imagined perspective-taking task after learning the two spatial layouts shown in Figure 1. Perspective-taking trials tested objects from layout one (left panel) or from layout two (right panel).

The interaction between layout one learning condition and layout two imagined perspective shows that previous experiences with layout one influenced the reference frame used to remember layout two. According to our interpretation, participants established a reference frame from the layout one learning perspective, and used that reference frame to represent the layout two locations learned later. Participants who learned layout one from 0° used a reference frame aligned with the learning perspective and the bilateral symmetry axis of the layout. When viewing layout two from 135°, they interpreted the layout two objects in the same reference frame used to represent layout one. Facilitation when imagining the 135° perspective of layout two suggests participants may have also represented layout two using a second reference frame aligned with the layout two study view. Participants who learned layout one from 135° represented layout one relative to a reference frame aligned with the 135° learning view. When viewing layout two from 135°, they interpreted those new objects in the same reference frame used to represent layout one. These findings specify a role for spatial reference frames during the development of spatial memories. However, it is unknown whether these findings will generalize to objects contained in different environments (e.g., neighboring rooms) or whether they are specific to objects contained within the same environment.

4 Transfer of Reference Frames across Sensory Modalities

The findings presented in Section 3 indicate that reference frames play an important role during the acquisition and development of spatial memories. Although vision is the primary sensory modality for acquiring spatial information, other senses also contribute to spatial learning. Research on learning through non-visual sensory modalities such as touch, proprioception, audition, and spatial language indicates that reference frames are also used to organize spatial memories learned through non-visual inputs (Avraamides, 2003; Avraamides & Kelly, in press; Yamamoto & Shelton, 2005, 2009; Yamamoto & Philbeck, 2008). However, it is unclear whether the *same* reference frames organize spatial memories learned through different sensory modalities. If so, then reference frames for locations learned through one sensory modality might influence learning through another sensory modality.

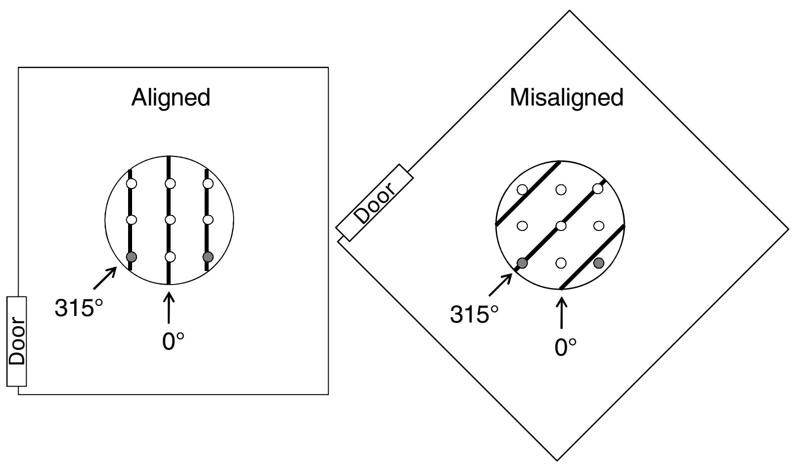


Fig. 3. Filled circles indicate locations of objects learned through vision; open circles indicate locations of objects learned only through touch. Objects were placed on a circular table with three stripes painted on its surface. Participants studied the two visual objects first from the 0° position and then from the 315° position. They then studied the touched objects only from the 315° position. The visual reference frame defined by the room walls and the stripes on the table were either aligned with the object grid (left panel, aligned condition) or misaligned with the object grid (right panel, misaligned condition).

To test whether reference frames exert cross-sensory influence, an experiment was conducted in which participants experienced salient visual cues emphasizing a reference frame before learning object locations through touch. The visual reference frame was defined by three bold stripes drawn on a round table (54 cm in diameter) on which the to-be-learned objects would be placed. The stripes were aligned with one axis of the surrounding rectangular room, providing a consistent visual reference frame. Participants visually studied two objects placed on the striped table (filled circles in Figure 3). The two visual objects were intended to direct participants' attention toward the cues defining the visual reference frame (especially the stripes on the table), and were never used during the perspective-taking task. Participants studied the visual objects from two views: first from the 0° view and then from the 315° view (participants were led 45° clockwise to reach the 315° view). After studying from the second view, participants were blindfolded and seven new objects were added to the table (open circles in Figure 3). Participants then studied all nine objects (the two objects previously studied visually plus the seven new objects) by touching them while seated at the 315° perspective. The nine objects formed a 3×3 grid with axes along 0°-180° and 90°-270°. In the aligned condition (Figure 3, left panel), the visual reference frame was aligned with the layout axes. In the misaligned condition (Figure 3, right panel), the visual reference frame was misaligned with the layout axes. Importantly, participants' egocentric experiences with the touched objects were exactly the same in both conditions, and the visual reference frame cues were blocked by a blindfold worn during the haptic learning phase.

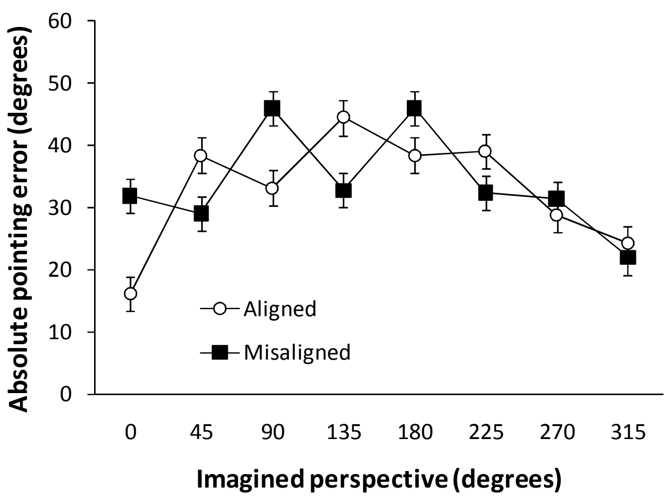


Fig. 4. Absolute pointing error during an imagined perspective-taking task using locations learned through touch (as shown in Figure 3). Participants studied the touched objects after experiencing a visual reference frame that was either aligned or misaligned with the axes of the layout of touched objects.

Subsequent perspective-taking trials used only the seven objects learned exclusively through touch. The visual objects were not used during perspective-taking. Performance is shown in Figure 4 as a function of imagined perspective for the aligned and misaligned conditions. Participants in the aligned condition performed best when imagining the 0° perspective compared to all other perspectives. Furthermore, the sawtooth pattern indicates that they remembered the touched objects using a reference frame composed of orthogonal axes along 0°-180° and 90°-270°. Participants in the misaligned condition performed best when imagining the 315° perspective compared to all other perspectives. In this case the sawtooth pattern suggests that they used a reference frame with orthogonal axes along 135°-315° and 45°-335°. In both conditions, performance from perspectives aligned with the visual reference frame may have been further enhanced by participants' alignment with the walls of the test room during the perspective taking task, similar to accounts of sensorimotor interference/facilitation (Kelly et al., 2007; May, 1996, 2004) The interaction between condition and imagined perspective provides clear evidence that the visual reference frame influenced spatial memories of subsequently touched objects. This indicates that reference frames exert their influence across sensory modalities, consistent with the hypothesis that reference frames are amodal.

5 Summary and Conclusions

Previous research has established the importance of reference frames in spatial memory organization, and has identified how environment-based and experience-based cues influence reference frame selection (e.g., Montello, 1991; Mou & McNamara, 2002; Shelton & McNamara, 2001). Here we extend those findings by describing a role for reference frames during the learning process, when learning occurs through one or more sensory modalities.

Section 3 described a new experiment showing that reference frames are selected when learning one part of an environment and then applied to other parts of that environment learned from different perspectives. Rather than using different reference frames to represent locations learned from different views, participants used a single reference frame to represent all objects in the environment. Once a reference frame had been established from the first view, that reference frame provided scaffolding for interpreting the locations of new objects learned from a different view. Learning in that experiment occurred within a circular room intended to limit the number of potential cues influencing reference frame selection. Future work should determine whether reference frames play a similar role when learning more complex and more natural environments.

Section 4 described a new experiment showing that reference frames exert their influence across sensory modalities. After experiencing a salient visual reference frame, participants used that same reference frame to organize locations of objects learned solely through touch. Similar to the experiment described in Section 3, participants used a single reference frame to represent the entire environment, rather than using separate reference frames to represent parts of the of the environment experienced visually and haptically. One possible mechanism for cross-modal reference frame transfer is that the visual reference frame might have influenced hand movements during subsequent haptic learning, such that the directions of hand movements were consistent with the visual reference frame (e.g., Lederman & Klatzky, 1987), but future research is needed to evaluate this hypothesis. Furthermore, it is unknown whether reference frames acquired through haptic experiences can influence locations subsequently learned through vision, and this is an important research question to help determine whether reference frames are truly amodal.

The present results may at first seem at odds with recent findings reported by Greenauer and Waller (2010). In their study, participants visually studied two adjacent object layouts containing bilateral symmetry axes that were misaligned with the learning view and also misaligned with each other. After learning, participants performed within- and between-layout judgments of relative direction. For both layouts, performance on within-layout judgments was best when imagining perspectives aligned with a layout's bilateral symmetry axis. In contrast, between-layout judgments were best from imagined perspectives aligned with the learning view or with the global geometry of the two layouts. Importantly, these findings occurred even when the layouts were learned sequentially. Overall, Greenauer and Waller's findings build on previous reports of hierarchically organized spatial memories (Hirtle & Jonides, 1985; Maki, 1981; McNamara, 1986; McNamara, Hardy & Hirtle, 1989; Stevens & Coupe, 1978) by showing that the hierarchical organization contains distinct micro-reference frames used to organize locations within spatial layouts and macro-reference frames used to organize spatial relations between layouts. In the current experiments, participants did not establish distinct reference frames to maintain the two sequentially-presented layouts. Instead, they established a single reference frame when viewing the first layout and used it to organize the second layout as well.

An important difference between the current experiments and those of Greenauer and Waller (2010) is that the two spatial layouts in the current experiments overlapped within the same space. Those overlapping layouts may have caused participants to organize their spatial memories around a single global reference frame rather than distinct local reference frames. Alternatively, the sequential presentation of one array embedded within another may have led participants to perceive the second array as a subset of the larger configuration. In this case, the current findings are compatible with those of Greenauer and Waller, showing that a macro-reference frame established for a superordinate layout can influence the selection of a micro-reference frame for a subordinate layout. Nevertheless, future research is needed to explore the conditions under which sequential spatial layouts become integrated or are organized separately in spatial memory.

Previous work has demonstrated that spatial memories are organized by reference frames centered on the environment. The research presented here defines a role for reference frames during the spatial learning process, when parts of the environment are encountered from different perspectives and through different sensory modalities. Reference frames provide an organizational structure for previously learned locations and for new locations not previously encountered. Reference frames influence spatial learning across sensory modalities, but further work is needed to determine whether modality-specific reference frames influence spatial learning through other sensory modalities or whether reference frames are truly amodal.

References

1. Avraamides, M.N.: Spatial updating of environments described in texts. Cognitive Psychology 47(4), 402–431 (2003)
2. Avraamides, M.N., Kelly, J.W.: Imagined perspective-changing within and across novel environments. In: Freksa, C., Nebel, B., Knauff, M., Krieg-Brückner, B. (eds.) Spatial Cognition IV. LNCS (LNAI), vol. 3343, pp. 245–258. Springer, Heidelberg (2005)

3. Avraamides, M.N., Kelly, J.W.: Multiple systems of spatial memory: Evidence from described scenes. Journal of Experimental Psychology: Learning, Memory & Cognition (in press)
4. Avraamides, M.N., Kelly, J.W.: Multiple systems of spatial memory and action. Cognitive Processing 9, 93–106 (2008)
5. Brockmole, J.R., Wang, R.F.: Changing perspective within and across environments. Cognition 87, B59–B67 (2003)
6. Golledge, R.G., Ruggles, A.J., Pellegrino, J.W., Gale, N.D.: Integrating route knowledge in an unfamiliar neighborhood: Along and across route experiments. Journal of Environmental Psychology 13, 293–307 (1993)
7. Greenauer, N., Waller, D.: Micro- and macro-reference frames: Specifying the relations between spatial categories in memory. Journal of Experimental Psychology: Learning, Memory, & Cognition (in press)
8. Hanley, G.L., Levine, M.: Spatial problem solving: The integration of independently learned cognitive maps. Memory & Cognition 11(4), 415–422 (1983)
9. Hintzman, D.L., O'Dell, C.S., Arndt, D.R.: Orientation in cognitive maps. Cognitive Psychology 13, 149–206 (1981)
10. Hirtle, S.C., Jonides, J.: Evidence of hierarchies in cognitive maps. Memory & Cognition 13, 208–217 (1985)
11. Holding, C.S., Holding, D.H.: Acquisition of route network knowledge by males and females. Journal of General Psychology 116, 29–41 (1988)
12. Ishikawa, T., Montello, D.R.: Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. Cognitive Psychology 52(2), 93–129 (2006)
13. Kelly, J.W., Avraamides, M.N., Loomis, J.M.: Sensorimotor alignment effects in the learning environment and in novel environments. Journal of Experimental Psychology: Learning, Memory & Cognition 33(6), 1092–1107 (2007)
14. Kelly, J.W., McNamara, T.P.: Spatial memories of virtual environments: How egocentric experience, intrinsic structure, and extrinsic structure interact. Psychonomic Bulletin & Review 15(2), 322–327 (2008)
15. Lederman, S.J., Klatzky, R.L.: Hand movements: A window into haptic object recognition. Cognitive Psychology 19, 342–368 (1987)
16. Maguire, E.A., Burke, T., Phillips, J., Staunton, H.: Topographical disorientation following unilateral temporal lobe lesions in humans. Neuropsychologia 34(10), 993–1001 (1996)
17. Maki, R.: Categorization and distance effects with spatial linear orders. Journal of Experimental Psychology: Human Learning & Memory 7, 15–32 (1981)
18. May, M.: Cognitive and embodied modes of spatial imagery. Psychologische Beitrage 38, 418–434 (1996)
19. May, M.: Imaginal perspective switchers in remembered environments: Transformation versus interference accounts. Cognitive Psychology 48, 163–206 (2004)
20. McNamara, T.P.: Mental representations of spatial relations. Cognitive Psychology 18, 87–121 (1986)
21. McNamara, T.P.: How are the locations of objects in the environment represented in memory? In: Freksa, C., Brauer, W., Habel, C., Wender, K.F. (eds.) Spatial Cognition III. LNCS (LNAI), vol. 2685, pp. 174–191. Springer, Heidelberg (2003)
22. McNamara, T.P., Halpin, J.A., Hardy, J.K.: Spatial and temporal contributions to the structure of spatial memory. Journal of Experimental Psychology: Learning, Memory, & Cognition 18(3), 555–564 (1992)

23. McNamara, T.P., Hardy, J.K., Hirtle, S.C.: Subjective hierarchies in spatial memory. Journal of Experimental Psychology: Learning, Memory & Cognition 15, 211–227 (1989)
24. McNamara, T.P., Rump, B., Werner, S.: Egocentric and geocentric frames of reference in memory of large-scale space. Psychonomic Bulletin & Review 10(3), 589–595 (2003)
25. Moar, I., Carleton, L.R.: Memory for routes. Quarterly Journal of Experimental Psychology A 34, 381–394 (1982)
26. Montello, D.R.: Spatial orientation and the angularity of urban routes: A field study. Environment and Behavior 23(1), 47–69 (1991)
27. Montello, D.R., Pick, H.L.: Integrating knowledge of vertically-aligned large-scale spaces. Environment and Behavior 25, 457–484 (1993)
28. Mou, W., McNamara, T.P.: Intrinsic frames of reference in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 28(1), 162–170 (2002)
29. Shelton, A.L., McNamara, T.P.: Systems of spatial reference in human memory. Cognitive Psychology 43(4), 274–310 (2001)
30. Shepard, R., Metzler, J.: Mental rotation of three dimensional objects. Science 171(972), 701–703
31. Stevens, A., Coupe, P.: Distortions in judged spatial relations. Cognitive Psychology 10, 422–437 (1978)
32. Valiquette, C.M., McNamara, T.P., Smith, K.: Locomotion, incidental learning, and the selection of spatial reference systems. Memory & Cognition 31, 479–489 (2003)
33. Waller, D., Lippa, Y., Richardson, A.: Isolating observer-based reference directions in human spatial memory: Head, body, and the self-to-array axis. Cognition 106, 157–183 (2008)
34. Waller, D., Montello, D., Richardson, A.E., Hegarty, M.: Orientation specificity and spatial updating of memories for layouts. Journal of Experimental Psychology: Learning, Memory, & Cognition 28, 1051–1063 (2002)
35. Wang, R.F., Spelke, E.S.: Updating egocentric representations in human navigation. Cognition 77, 215–250 (2000)
36. Wang, R.F., Spelke, E.S.: Human spatial representation: insights from animals. Trends in Cognitive Sciences 6(9), 376–382 (2002)
37. Werner, S., Schmidt, K.: Environmental reference systems for large-scale spaces. Spatial Cognition and Computation 1(4), 447–473 (1999)
38. Yamamoto, N., Philbeck, J.W.: Egocentric and intrinsic frames of reference in haptic spatial learning. Poster presented at the 49th annual meeting of the Psychonomic Society, Chicago, IL (2008)
39. Yamamoto, N., Shelton, A.L.: Visual and proprioceptive representations in spatial memory. Memory & Cognition 33, 140–150 (2005)
40. Yamamoto, N., Shelton, A.L.: Orientation dependence of spatial memory acquired from auditory experience. Psychonomic Bulletin & Review 16(2), 301–305 (2009)

Do We Need to Walk for Effective Virtual Reality Navigation? Physical Rotations Alone May Suffice*

Bernhard E. Riecke[1,2], Bobby Bodenheimer[1], Timothy P. McNamara[1], Betsy Williams[1,3], Peng Peng[1], and Daniel Feuereissen[1,2]

[1] Vanderbilt University, Nashville, TN, USA
ber1@sfu.ca
[2] Simon Fraser University, Surrey, BC, Canada
[3] Rhodes College, Memphis, TN, USA

Abstract. Physical rotations and translations are the basic constituents of navigation behavior, yet there is mixed evidence about their relative importance for complex navigation in virtual reality (VR). In the present experiment, 24 participants wore head-mounted displays and performed navigational search tasks with rotations/translations controlled by physical motion or joystick. As expected, physical walking showed performance benefits over joystick navigation. Controlling translations via joystick and rotations via physical rotations led to better performance than joystick navigation, and yielded almost comparable performance to actual walking in terms of search efficiency and time. Walking resulted, however, in increased viewpoint changes and shorter navigation paths, suggesting a rotation/translation tradeoff and different navigation strategies. While previous studies have emphasized the importance of full physical motion via walking (Ruddle & Lessels, 2006, 2009), our data suggests that considerable navigation improvements can already be gained by allowing for full-body rotations, without the considerable cost, space, tracking, and safety requirements of free-space walking setups.

1 Introduction

Virtual reality and other multi-media technologies that can create computer-mediated experiences have become more widespread, affordable, and accepted in recent years. In fact, as a large body of fiction literature and cinema have shown, there have been many exaggerated claims about the potential experience virtual reality will ultimately offer – that is, an almost super–human experience that far exceeds real-world possibilities (see, e.g., movies like *The Matrix*). The level of realism and detail of current audio-visual simulations can indeed be stunning. Despite those promising claims and achievements, however, the user experience of virtual reality (VR) does not match its fictional characterization, and there remain a number of major challenges. One of these challenges is spatial orientation. Even though one might hope that VR should ultimately enable us to locomote and orient as well as in the real world – or potentially even better, as VR

* We thank John Rieser for fruitful discussions. Support: Vanderbilt University; NIMH grant 2-R01-MH57868; P30-EY008126; NSF 0821640; NSERC 371799; and Simon Fraser University.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 234–247, 2010.

can allow for the physically impossible (e.g., teleporting, flying, bird eye's views, or multiple and even simultaneous perspectives to chose from) – spatial orientation in VR can be quite poor.

1.1 Spatial Orientation in VR

Under certain circumstances VR-users take significantly longer to learn virtual environments than comparable real environments (Richardson, Montello, & Hegarty, 1999; Witmer, Bailey, Knerr, & Parsons, 1996), and often produce large random and systematic errors in virtual environment (VE) navigation (Riecke, 2008). Recent studies have demonstrated that participants in visually-based VR also produce certain novel types of qualitative errors such as left-right confusion or failure to update visually simulated rotations altogether (Klatzky, Loomis, Beall, Chance, & Golledge, 1998; Avraamides, Klatzky, Loomis, & Golledge, 2004; Riecke, 2008). Note that such qualitative errors do not occur in comparable real-world situations where participants are allowed to physically walk the trajectory that was only visually simulated in the VR task (Klatzky et al., 1998). Furthermore, retrieval of memories of a real environment seems to be affected by features of the environment (Riecke & McNamara, 2007) in a way that retrieval of memories of virtual environments is not (Moura & Riecke, 2009; Williams, Narasimham, Westerman, Rieser, & Bodenheimer, 2007). In summary, there seem to be many situations where spatial orientation performance in VR is worse than real-world performance.

There are, however, a few noteworthy exceptions where visual cues presented in immersive VR can be sufficient for spatial orientation tasks: When high visual realism is combined with an abundance of reliable landmarks and an immersive HMD or projection system, visual cues may be sufficient to enable excellent homing performance (Riecke, van Veen, & Bülthoff, 2002) as well as spatial updating performance that approaches real-world performance (Riecke, von der Heyde, & Bülthoff, 2005).

In general, though, performance in VR spatial orientation tasks seems to benefit from physical locomotion in the environment (Avraamides et al., 2004; Chance, Gaunet, Beall, & Loomis, 1998; Klatzky et al., 1998; Pausch, Proffitt, & Williams, 1997; Ruddle & Lessels, 2006; Waller, Beall, & Loomis, 2004; Wraga, Creem-Regehr, & Proffitt, 2004). Especially when participants are asked to respond not only as accurately as possible but also as fast as possible, visual path-integration based spatial orientation in VR often shows strikingly large systematic as well as random errors (Riecke, 2008). In sum, despite impressive advances in VR hardware and software, human spatial orientation is often still far from the real-world-like performance, where automatic spatial updating can allow for intuitive yet effective spatial orientation with minimal cognitive load (Farrell & Robertson, 1998; Klatzky et al., 1998, e.g.,). The current study was designed to elucidate the origins of this difference in human navigation ability between real and virtual environment by assessing the relative contribution and relevance of physical translations and rotations.

1.2 Goal of This Study: Investigate Relative Contribution of Physical Translations and Rotations for Effective Navigation in VR

Whereas self-motions in VR can be easily visually simulated using affordable off-the-shelf hardware and software, creating a VR system that allows users to physically walk in a natural manner through the simulated world requires considerable cost, effort, and safety measures. Such a system typically requires a sufficiently large obstacle-free space for participants to walk unimpeded, the use of head-mounted displays (HMD), the ability to track the motion of the HMD, and an experimenter present to ensure that the participant does not walk into physical walls, stumble over cables, etc. It is therefore critical to assess if there is sufficient benefit from physical locomotion/walking to justify the high effort and cost associated with large free-space walking – especially given that physical rotations alone can be realized without the need for large tracked free-space walking areas. In this context, a series of studies by Ruddle & Lessels (2006, 2009) is of particular importance and will be discussed here in more detail, as it addresses these issues directly and inspired the experimental paradigm of the current study.

1.3 Navigational Search Studies by Ruddle and Lessels

Ruddle and Lessels performed a series of navigational search studies in which participants were in a real or virtual rectangular room that contained a regular arrangement of 32 pedestals, half of which had closed boxes placed on top (Lessels & Ruddle, 2005; Ruddle & Lessels, 2006, 2009). Participants were asked to search for eight target objects hidden in these 16 boxes.

In a real-world test, participants performed with almost perfect efficiency: participants found all eight targets without any re-visits of boxes in 93% of trials (Lessels & Ruddle, 2005). Even when blinders restricted the field of view to just $20°\times16°$, performance was not significantly reduced (87% perfect trials without revisits, see Fig. 2). When performing a similar task using an HMD with a FOV of $48°\times36°$ and physical walking, performance was similar (90% trials without revisits), and was independent of whether the visual scene was modeled with high or low visual realism (Ruddle & Lessels, 2006). This suggests that free-space walking with an HMD can allow for performances that matches real-world performance, at least for the task at hand. However, performance was substantially reduced when visually simulated translations were controlled using a button-press interface ("real rotation" condition) instead of physical walking (43% no-revisit trials, see Fig. 2), even though rotations in the VE were still controlled by physical rotations (Ruddle & Lessels, 2006). This suggests that both translational and rotational physical motion cues are required to allow for real-world-like performance in such search tasks.

When the HMD was replaced by a monitor and rotations and translations were controlled by moving a computer mouse and pressing buttons on a keyboards, respectively (Ruddle & Lessels, 2006), performance did not significantly decrease further (45% no-revisit trials in the "visual only" condition, see Fig. 2). One might be tempted to conclude from those data that adding physical rotations did not show any performance benefit (and the authors do, in fact, suggest that). It is possible, however, that the experimental methodologies in (Ruddle & Lessels, 2006, 2009) could have contributed to the observed similarity in performance for the visual-only and real rotation condition:

Whereas Ruddle & Lessels (2006, 2009) used the same display (a stereo HMD with 640×480 resolution) for the walking condition and real rotation/keyboard translation condition, a different display device (a 21 inch desktop monitor) was used in the visual only (mouse rotation/keyboard translation) condition. This change in display device produced several differences in the nature of the visual experiences that might have contributed to Ruddle and Lessels's findings. Potential differences include display quality and resolution (presumably higher for monitor); availability of binocular depth cues (HMD only); tracking latency (HMD only); peripheral visibility of surrounding room (monitor only); immersion and presence (higher perhaps with HMD); eye height (sitting for monitor vs. standing for HMD); mismatch between simulated eye height (always 1.65m) and physical eye height (especially in monitor condition); and absolute size of the display (21" for monitor vs. 2×1.3" for HMD).

Many of these parameters have been shown to impact human spatial orientation performance (Riecke, Schulte-Pelkum, & Bülthoff, 2005; Riecke et al., 2002; Tan, Gergle, Scupelli, & Pausch, 2006, 2004; Tan, Gergle, Scupelli, & R.Pausch, 2005; Alfano & Michel, 1990). Moreover, Ruddle and colleagues themselves had demonstrated that replacing an HMD with a desk-top monitor reduces navigation performance in VR, indicated by increased navigation times and less accurate sense of straight-line distances (Ruddle, Payne, & Jones, 1999). Hence, any direct comparison between results from the monitor (visual only) condition and the HMD (real rotation and walking) conditions should be treated with caution. The current study was designed to avoid these potential problems by using the same HMD in all conditions, thus allowing for a direct and more trustworthy comparison between all three conditions, and thus a more reliable assessment of the relative importance and contribution of physical translations versus rotations for VR navigation. To this end, we used a similar overall navigational search methodology as Ruddle & Lessels (2006, 2009), with the following major changes:

– **Same Display for all Conditions.** The same visual display (HMD) was used in all conditions.
– **Orientating Cues from Environmental Geometry and Rectangular Object Structure Removed.** Recently, Kelly, McNamara, Bodenheimer, Carr, & Rieser (2008) showed that navigators can use room geometry to maintain orientation during spatial updating in immersive VR. This suggests that participants in Ruddle & Lessels (2006, 2009) might have used the geometry of the surrounding room as well as the rectangular arrangement and regular orientation of the objects to reorient themselves or maintain global orientation, which might in turn have contributed to the observed lack of any benefit of physical rotations (as visual cues were sufficient for effective (re)orientation). To avoid any influence of environmental geometry or intrinsic reference frame of the object layout, we removed the surrounding room and randomly positioned and oriented all objects for each trial. Furthermore, we refrained from using a naturalistic, landmark-rich environment, as this might have obfuscated potential effects of locomotion mode due to, for example, instantaneous spatial updating or other kinds of powerful visual re-orientation mechanisms (Riecke et al., 2005; Riecke, Cunningham, & Bülthoff, 2007).
– **Joystick as Continuous Input Device.** While Ruddle & Lessels (2006, 2009) used a simple button-based translation input paradigm in the visual only and real rotation condition where participants could not continuously adjust their translational velocity, we used a wireless joystick that allowed for easy and continuous adjustment of translational and rotational velocities.

- **Posture and Eye Height Matched for all Conditions.** Participants were always standing, with the simulated eye height matching their actual eye height.
- **Within-Participant Design.** As between-participant variability is often strikingly large for spatial orientation tasks (even more so in VR), we used a complete within-participant design for the current study.

2 Methods

Participants. Twenty-four adults (19–46 years old, half female) from the Nashville community participated for monetary compensation.

2.1 Stimuli and Apparatus

Virtual Environment. Participants wore an HMD that displayed a simple virtual scene that contained a large textured ground plane and 16 identical objects (randomly oriented 3D models of a pedestal with a miniature birdhouse on top). Eight of the 16 birdhouses contained red balls as target objects that participants had to search for, and the remaining eight birdhouses were empty and acted as decoys. The position of the targets objects was randomly Poisson-disk distributed within a circular area of 4m diameter for each trial to avoid orienting cues and learning effects. This size was chosen such that the object layout fitted safely withing the 5 × 5m free space walking area of the actual lab. As we were interested in investigating different locomotion modes, we did not simulate room geometry or other potential landmarks or regular object layouts that could have been used to guide or re-orient participants (Kelly et al., 2008). In particular, the ground texture and sky contained no orientation cues.

Visualization. The virtual scene was presented binocularly through a head-mounted display (NVIS NVisor SX) at a native resolution of 1280×1024 pixel with complete binocular overlap. The simulated field of view (FOV) matched the physical FOV of 60° diagonal (corresponding to 48°×36° or 27 pixel/°).

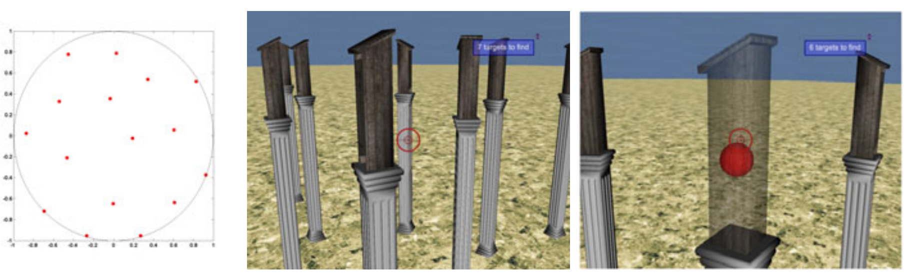


Fig. 1. Left: Sample unit circle Poisson-disk distribution of the pedestal array, which was randomized per trial. These values were scaled by the experimental radius (2m) and represented the positions of the pedestals on the ground plane. **Middle**: Experimental view, with the cross-hair indicating the current motion direction. **Right**: By pressing a designated button on the joystick, participants could look inside a birdhouse for 2 seconds to see if it contained a target object (red ball).

Tracking and Interaction. The HMD was tracked at 60Hz using the WorldViz PPT tracking system. Participants carried a wireless joystick (Logitech Freedom 2.4) that was mounted on a wooden board that was worn using shoulder straps.

Headphones and Ambient Sound Masking. To mask spatialized auditory cues that could have been used to reorient in the lab, a broad-frequency river-like sound was displayed via the headphones integrated in the HMD throughout the experiment.

Task. Participants' task was to navigate the VE until they had encountered each of the eight targets, using a designated button on the joystick to render the walls of the birdhouses semi-transparent for two seconds. Additional auditory cues indicated upon button-press whether the object was a target or decoy. Birdhouses that had already been visited were populated with blue balls: either a blue ball appeared if the birdhouse was previously empty, or the red ball turned blue upon visitation. If targets were revisited, a different sound was also played and a blue ball was visible. Participants were instructed to minimize the number of re-visits, distance traveled, and time needed without starting to rush or run through the VE. A trial was terminated when either all eight target objects were found or when there were eight consecutive revisits (i.e., looking into previously-visited birdhouses). A message in the upper right hand corner of the screen displayed the current number of targets left to find in a trial.

Locomotion Modes. To investigate the relative contribution and importance of physical motion cues for translation and rotation, we employed three different locomotion methods:

In the "**walking**" condition, all six degrees of freedom of the HMD were tracked and participants navigated through the virtual scene by physically walking (translating and turning).

In the "**real rotation**" condition, tracking of physical translations in the horizontal plane was switched off, and participants instead had to use joystick deflections to translate through the virtual scene. Rotations and up-down translations were still controlled by corresponding physical motions.

In the "**joystick**" condition, both horizontal translations and yaw/pitch rotations in the VE were controlled by the joystick deflections and rotations, respectively.

Note that participants were standing in all motion conditions.

Motion Model. Joystick deflections were linearly mapped onto translational velocities in the VE ("velocity control" mapping). Yaw rotations were controlled by yaw-rotations of the joystick handle using a velocity control mapping, whereas pitch rotations were controlled using the lever on the side of the joystick using a position control mapping (as that lever did not have a re-centering force). Maximum joystick rotations and deflections resulted in velocities that matched the typical maximum physical motion velocities we observed during pre-tests in the full-motion condition (1m/s and 90°/s). A joystick dead zone was used to enable precise navigation and to ensure that participants were stationary in the VE when they did not touch the joystick.

2.2 Procedure and Experimental Design

Each participant completed one practice trial and three test trials for each of the three locomotion modes, resulting in a total of twelve trials[1]. Locomotion modes were blocked and balanced across participants. Trials lasted 2min on average, and participants took anywhere between 40s and 4min to complete a given trial. To reduce the likelihood of fatigue and simulator sickness, participants were allowed to take off the HMD and take a short break after each trial. Additional breaks were scheduled after each block of four trials. The pedestal layout and orientation were independently randomized for each trial to prevent layout learning. Participants started each trial at the perimeter of the object array facing inwards. After the experiment, participants were debriefed, payed, and thanked for their participation.

3 Results

Data for the different dependent measures are summarized in Fig. 3 and analyzed using repeated-measures ANOVAs and planned pairwise contrasts for the independent variable locomotion mode (3 levels: walking; real rotation and joystick-based translation; joystick rotation and translation). Statistical results are summarized in Table 1.

Number of Perfect Trials Lower Than in Ruddle and Lessels (2006) Study. Comparing overall performance between the current study and the study by Ruddle & Lessels (2006) shows two main differences, as illustrated in Fig. 2: Whereas participants in Ruddle & Lessels (2006) were able navigate "perfectly" (i.e., find all targets without any revisits) in 90% of the trials in the HMD walking condition, performance dropped to only 43% and 45% perfect trials in the real rotation and visual only condition, respectively (cf. Fig. 2). In contrast, participants in the current study performed overall considerably worse, irrespective of the motion condition (cf. Fig. 2): Only 13.9% of the trials were finished without any revisits in the walking condition, and performance did not significantly drop further when physical motion cues were excluded for translations (real rotation condition) and rotations (joystick condition) (cf. Table 1).

Number of Revisits Needed to Complete the Task Was Highest for Joystick Condition. Whereas participants in the joystick condition revisited on average 6.4 pedestals before completion, being able to walk significantly reduced this number to only 4.2 revisits (cf. Fig. 3a and Table 1). This might be interpreted a participants being less lost in the walking condition. Interestingly, the contrast analysis revealed that replacing walking with joystick translations in the real rotation condition did not significantly impair performance, whereas replacing actual rotations with visually simulated rotations in the joystick condition marginally increased the number of revisits. A similar data pattern was found for the number of targets revisited (Fig. 3b): Whereas the joystick condition yielded significantly more revisits than the walking condition, performance in the real rotation condition equaled the walking condition.

[1] Twelve additional trials that used a gain factor of 10:1 between visually simulated and actual translations were excluded due to technical problems with the tracking. These 10:1-gain trials were performed in one session, either before or after the 1:1 gain trials. The order of the 10:1 versus 1:1-gain sessions was balanced to avoid systematic order effects.

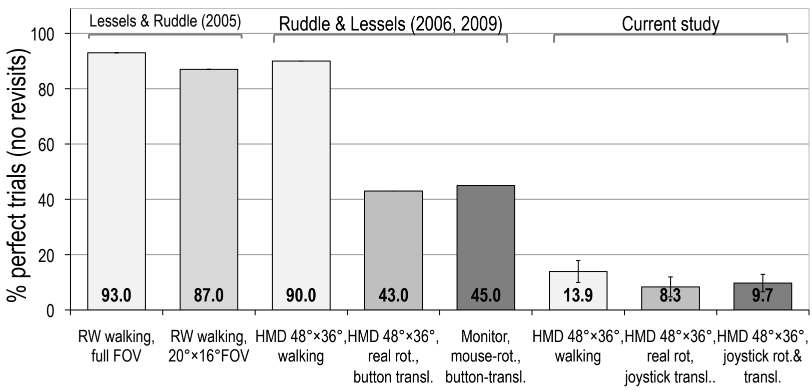


Fig. 2. Re-plotting of the data from Lessels & Ruddle (2005) and Ruddle & Lessels (2006, 2009) for comparison with the current data

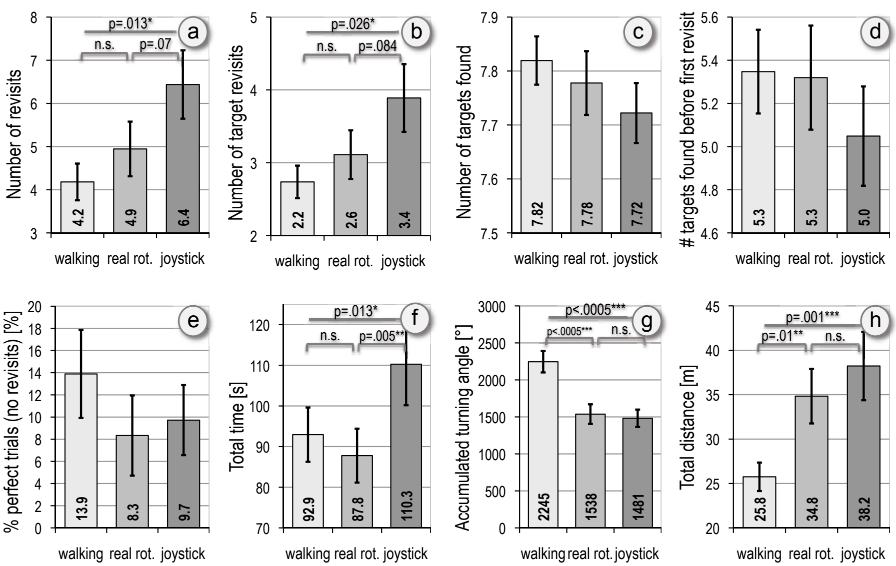


Fig. 3. Mean data for the different dependent measures. Error bars indicate ±1 standard error of the mean.

Number of Targets Found Did Not Depend on Motion Condition. Figure 3d shows that participants on average found about five targets before their first unplanned revisit of any pedestal. This measure ("number of targets found before first revisit") could thus be taken as a conservative estimate of when on average participants first got "lost" in the environment and could not tell any more whether they had previously visited a given location or not. The lack of any significant differences between the motion conditions

Table 1. Analysis of variance and planned pairwise contrasts results for the different dependent variables. The asterisks indicate the significance level (α = 5%, 1%, or 0.1%). Marginally significant effects ($\alpha \leq$ 10%) are indicated by an *'m'*. Significant and marginally significant effects are typeset in bold and italics, respectively. The effect strengths partial η_p^2 indicates the percentage of variance explained by a given factor.

	ANOVA main effect			Contrast real walking vs. real rotation			Contrast real rotation vs. joystick		
	F(2,46)	p	η_p^2	F(1,23)	p	η_p^2	F(1,23)	p	η_p^2
Number of revisits	**4.803**	**.013****	**.173**	1.211	.282	.050	*3.610*	*.070m*	*.136*
Number of target revisits	**3.858**	**.026****	**.147**	.983	.332	.041	*3.262*	*.084m*	*.124*
Number of targets found	1.060	.355	.044	.282	.601	.012	.885	.357	.037
# targets found before 1st revisit	.608	.511	.026	.007	.934	.000	.637	.433	.027
% perfect trials (no revisits)	.926	.404	.039	1.353	.257	.056	.138	.714	.006
Total time	**4.814**	**.013***	**.173**	.560	.462	.024	**9.430**	**.005****	**.291**
Accumulated turning angle	**25.20**	**<.0005*****	**.523**	**29.47**	**<.0005*****	.562	.491	.490	.021
Total distance traveled	**7.621**	**.001*****	**.249**	**7.843**	**.010****	**.254**	1.130	.299	.047

(cf. Table 1) suggest that the availability of physical motion cues in the walking condition did not provide any useful cues that participants could have successfully used to remain oriented and prevent revisits. Note that not all trials were successfully completed (cf. Fig. 3c), as trials were automatically aborted after eight successive revisits to previously-visited pedestals. In the walking condition, an average of 7.82 out of 8 targets were found, and this number did not drop significantly for the real rotation or joystick condition (cf. Table 1).

Navigational Search Was Fastest in Walking and Real Rotation Condition. Figure 3f shows that participants took overall about 25% longer in the joystick condition, as compared to the real rotation condition where head motions instead of joystick motions were used to control orientation changes in VR. Allowing for actual walking did not reduce navigational search time further, suggesting that physical translations are less critical than physical rotations for search efficiency.

Walking Led to Increased Orienting Motions and Reduced Path Lengths. Figure 3h shows that participants covered about 35% less overall distance in the walk condition (25.8m) than in the real rotation condition (34.8m). This decrease in traveled distance for walking was, however, accompanied by a 46% increase in the amount of head rotations (Fig. 3g): Accumulated turning angles significantly increased from 1538° (or 4.27 revolutions) to 2245° (or 6.24 revolutions). This suggests a qualitatively different navigation strategy in the walking condition: By looking around more in the walking condition were participants able to optimize the trajectory, thus traveling less far. Note that this strategy did not, however, reduce the amount of time needed for the navigational search task in the walking condition as compared to the real rotation condition (Fig. 3f). In fact, apart from a reduction in the amount of distance walked, none of the seven other dependent measures shows any significant benefit of physically walking as compared to joystick translations combined with physical rotations.

4 Discussion and Conclusions

In an important series of studies on navigational search in real and virtual environments, Ruddle and Lessels found that participants performed better when allowed to freely walk, as compared to a "real rotation" condition where they wore an HMD and could freely rotate while controlling simulated translations using a button-based motion model (Lessels & Ruddle, 2005; Ruddle & Lessels, 2006, 2009). This real rotation condition led to similar performance as a visual only condition, where both translations and rotations were only visually simulated on a desktop monitor, and controlled via keyboard presses and mouse motions, respectively.

These data led Ruddle & Lessels (2006) to posit that full physical movement is essential for effective navigation in VR, whereas physical rotations alone are insufficient. The current study replicated the overall navigational search procedure of Ruddle & Lessels (2006) while controlling for a number of variables that could have affected the previous results and interpretations (see discussion in the introduction section). Most importantly, we used the same display device (HMD) for all motion conditions, and carefully removed salient landmarks and other potential visual orienting cues by removing the rectangular room environment and randomly positioning and orienting all objects in the scene for each trial. Note that these changes were expected to increase overall task difficulty. Furthermore, we used a joystick-based motion paradigm that allowed for smooth, continuous and intuitive adjustment of rotational and translational velocities instead of the button- and mouse-based motion model used in (Ruddle & Lessels, 2006, 2009). Controlling for these variables in the current study led to a qualitatively different pattern of results.

Whereas there was still an overall benefit for full physical motion (walking condition) as compared to joystick navigation, merely allowing for bodily rotations (real rotation condition) provided considerable performance benefits when compared to the joystick (visual only) condition. Moreover, real rotation performance almost equalled walking performance: Only one of the eight dependent measures (total distance traveled) showed a clear and significant benefit of walking over real rotation. In fact, participants in the walking condition of our study seemed to trade off distance for rotations, insofar as they walked less but turned more, which might be caused by a shift in navigation strategy.

In sum, comparing previous results by Ruddle & Lessels (2006, 2009) with our results raises several major questions that we will discuss in the following subsections.

Did the Lack of Orienting Cues in our Study Cause the Performance Drop Compared to Ruddle and Lessels (2006, 2009)? As the overall navigational search paradigm was quite similar in the current study and (Ruddle & Lessels, 2006, 2009) (especially for the walking condition), we propose that the large overall performance decrement in the current study might be caused by differences in environmental geometry and the intrinsic geometric structure of the object array: Ruddle & Lessels (2006) used a rectangular surrounding room and a rectangular grid-like structure of the pedestals with additional constraints (always one target and one decoy per group of four pedestals), whereas the current study did not contain any such regularities or other environmental geometry cues that participants could have used to remain oriented or re-orient. As the other procedural differences in the walking condition seem minimal, the large overall performance difference between our and Ruddle and Lessels's studies suggests that the rectangular structure of the room and object array layout was indeed an important factor for participants' ability to remain oriented in Ruddle & Lessels (2006), and could have served as a reference frame for remaining oriented in VR (McNamara, Sluzenski, & Rump, 2008). This hypothesis is in agreement with recent results by Kelly et al. (2008), who showed that the environmental geometry of the surrounding room in VR is an important factor for both spatial updating and reorientation. In particular, when the rectangular room was replaced by a cylindrical room devoid of orienting cues, participants got increasingly lost for increasing path lengths. Similarly, Riecke et al. (2005, 2007) showed that naturalistic environmental cues, but not optic flow alone, can be sufficient for spatial updating with performance approaching real-world performance, with or without concurrent physical motion cues.

Why Did Real Rotation Performance Approach Walking Performance in Our Study, But Not in Ruddle and Lessels (2006, 2009)? Ruddle & Lessels (2006) observed a clear performance benefit of walking over real rotation without physical translation, but the current study did not, despite using a similar navigational search task. Whereas the increase in task complexity due to the randomized target configuration and lack of environmental geometry in our study can explain the overall performance decrement, it seems unclear how task complexity could differentially affect performance in the walking versus real rotation condition. Hence, we propose that differences in the translational motion paradigm might be (at least in part) responsible for the observed differences: Ruddle & Lessels (2006) use a button-based translation control that allowed only forward motions and did not allow participants to continuously adjust their speed, whereas the current study used a joystick that allowed for continuous and intuitive control of translational velocities in the horizontal plane. Although further studies are needed to corroborate this conjecture, the current data highlights the importance of devising and carefully testing input-devices that allow for virtual locomotion with similar intuitive control, accuracy, and precision as actual walking while minimizing the cognitive load.

Why Did Real Rotation Performance Exceed Visual-Only Performance in our Study, But Not in Ruddle and Lessels (2006, 2009)? Previous studies suggest that adding physical rotations cues can improve spatial orientation performance compared to visual-only simulations for various basic spatial tasks (Bakker, Werkhoven, & Passenier, 1999; Klatzky et al., 1998; Lathrop & Kaiser, 2002; Pausch et al., 1997).

However, for more complex spatial orientation and navigation tasks, there seems no clear evidence that physical rotations themselves improve performance, although providing full physical motions did prove to be beneficial. (Chance et al., 1998; Ruddle & Lessels, 2006, 2009). When asked to point to previously-encountered targets in a HMD-based maze tasks, participants pointed more accurately when walking as compared to pure joystick navigation (Chance et al., 1998). A real rotation condition where translations were joystick-controlled and rotations physically controlled yielded intermediate performance, with no significant differences relative to either walking or joystick condition. As discussed earlier, navigational search performance in (Ruddle & Lessels, 2006, 2009) showed no benefit of adding physical rotations over visual-only navigation. This is in contrast to the current study, which used the same navigational search task as (Ruddle & Lessels, 2006, 2009) but showed a clear benefit for added physical rotations. There are a number of factors that could have contributed to the observed differences:

First, Ruddle & Lessels (2006, 2009) used different displays (HMD vs. monitor) for the real rotation and visual-only rotation condition, whereas the current study used the same HMD for all conditions. Second, the navigational search task was likely more difficult in our study, as salient (re-)orienting cues such as the rectangular environment and the regular, rectangular object layout in (Ruddle & Lessels, 2006, 2009) were replaced by randomized object positions and orientations, thus largely avoiding salient orienting cues and intrinsic/extrinsic reference frames that participants could have used to remain oriented or re-orient during their simulated movements (Kelly et al., 2008; Riecke et al., 2005). Although further careful experimentation is needed, the comparison suggests that the availability of ample orienting cues in (Ruddle & Lessels, 2006, 2009) might have obscured potential effects of physical rotations, and that physical rotation cues might become more important under high cognitive load/task difficulty and limited availability of visual (re-)orienting cues.

Did Participants in the Joystick Condition Perform Poorly Because they Failed to Update their Heading? Previous studies investigating the updating of cognitive heading in VR showed individual differences in participants' strategies and the resulting systematic errors and underlying neural representation (Gramann, Muller, Eick, & Schonebeck, 2005; Iaria, Petrides, Dagher, Pike, & Bohbot, 2003). For example, using a point-to-origin paradigm after a visually simulated 2-segment excursion displayed on a desktop monitor, Gramann et al. (2005) reported that more than half of their participants responded as if they had not updated their heading and were still facing the original orientation. Gramann et al. interpreted these group of participants as "non-turners" who might have used an allocentric strategy that did not incorporate visually simulated heading changes. Klatzky et al. (1998) had reported similar failures to incorporate heading changes that were not physically performed but only visually displayed (in an HMD condition) or verbally instructed (in an imagine condition). Together, these studies suggests that participants in the current study might have shown comparable failures to update heading changes in the joystick condition where the visually simulated rotations were not physically performed, which might have contributed to the reduced task performance. The current experiment was, however, not designed to investigate this issue and does not allow for any specific conclusions, although informal observations suggest that the low overall performance in the joystick condition often coincided with being lost and disoriented. Additional reference frame conflicts between participants' physical orientation in the lab and the simulated orientation in the joystick condition might have further contributed to the overall poor performance in the joystick condition (McNamara et al., 2008; Riecke, 2008).

Conclusions and Outlook. Whereas previous studies showed a clear benefit of physical rotation cues only for simple spatial task (Bakker et al., 1999; Klatzky et al., 1998; Lathrop & Kaiser, 2002; Pausch et al., 1997), but not for more complex navigation tasks (Chance et al., 1998; Ruddle & Lessels, 2006, 2009), the current study provides first evidence that allowing VR users control simulated rotations with their own body can have significant benefits over mere joystick navigation. Moreover, navigation performance in this real-turn mode was statistically equivalent to performance for actual walking in six out of eight dependent variables, and the real-turn mode even reduced the amount of viewing direction changes significantly. These results suggest that, for many applications, allowing for full-body rotations without actual walking can provide considerable performance benefits, even for complex and cognitively demanding navigation tasks. These findings can help to reduce overall simulation effort and cost, as allowing users to walk through VR requires sufficiently large, position-tracked free-space walking areas and additional safety measures, whereas physical rotations can be implemented more easily and cost-effectively.

References

Alfano, P.L., Michel, G.F.: Restricting the field of view: Perceptual and performance effects. Perceptual and Motor Skills 70, 35–45 (1990)

Avraamides, M.N., Klatzky, R.L., Loomis, J.M., Golledge, R.G.: Use of cognitive versus perceptual heading during imagined locomotion depends on the response mode. Psychological Science 15(6), 403–408 (2004)

Bakker, N.H., Werkhoven, P.J., Passenier, P.O.: The effects of proprioceptive and visual feedback on geographical orientation in virtual environments. Presence - Teleoperators and Virtual Environments 8(1), 36–53 (1999)

Chance, S.S., Gaunet, F., Beall, A.C., Loomis, J.M.: Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. Presence - Teleoperators and Virtual Environments 7(2), 168–178 (1998)

Farrell, M.J., Robertson, I.H.: Mental rotation and the automatic updating of body-centered spatial relationships. Journal of Experimental Psychology – Learning Memory and Cognition 24(1), 227–233 (1998)

Gramann, K., Muller, H.J., Eick, E.M., Schonebeck, B.: Evidence of Separable Spatial Representations in a Virtual Navigation task. Journal of Experimental Psychology – Human Perception and Performance 31(6), 1199–1223 (2005)

Iaria, G., Petrides, M., Dagher, A., Pike, B., Bohbot, V.D.: Cognitive Strategies Dependent on the Hippocampus and Caudate Nucleus in Human Navigation: Variability and Change with Practice. J. Neurosci. 23(13), 5945–5952 (2003)

Kelly, J.W., McNamara, T.P., Bodenheimer, B., Carr, T.H., Rieser, J.J.: The shape of human navigation: How environmental geometry is used in maintenance of spatial orientation. Cognition 109(2), 281–286 (2008)

Klatzky, R.L., Loomis, J.M., Beall, A.C., Chance, S.S., Golledge, R.G.: Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. Psychological Science 9(4), 293–298 (1998)

Lathrop, W.B., Kaiser, M.K.: Perceived Orientation in Physical and Virtual Environments: Changes in Perceived Orientation as a Function of Idiothetic Information available. Presence-Teleoperators and Virtual Environments 11(1), 19–32 (2002)

Lessels, S., Ruddle, R.A.: Movement around real and virtual cluttered environments. Presence: Teleoper. Virtual Environ. 14(5), 580–596 (2005)

McNamara, T., Sluzenski, J., Rump, B.: Human Spatial Memory and Navigation. In: Byrne, J.H. (ed.) Learning and Memory: A Comprehensive Reference, pp. 157–178. Academic Press, Oxford (2008)

Moura, D., Riecke, B.E.: Is Seeing a Virtual Environment Like Seeing the Real Thing? In: ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV), p. 131. ACM, New York (2009)

Pausch, R., Proffitt, D., Williams, G.: Quantifying immersion in virtual reality. In: SIGGRAPH 1997: Proceedings of the 24th annual conference on Computer graphics and interactive techniques, New York, NY, USA, pp. 13–18. ACM Press/Addison-Wesley Publishing Co. (1997)

Richardson, A.E., Montello, D.R., Hegarty, M.: Spatial knowledge acquisition from maps and from navigation in real and virtual environments. Mem. Cogn. 27(4), 741–750 (1999)

Riecke, B.E.: Consistent Left-Right Reversals for Visual Path Integration in Virtual Reality: More Than a Failure to Update One's Heading? Presence - Teleoperators and Virtual Environments 17(2), 143–175 (2008)

Riecke, B.E., McNamara, T.P.: Similarity Between Room Layouts Causes Orientation-Specific Sensorimotor Interference in To-Be-Imagined Perspective Switches. In: Proceedings of the 48th Annual Meeting of the Psychonomic Society (Psychonomics), p. 63 (2007)

Riecke, B.E., Schulte-Pelkum, J., Bülthoff, H.H.: Perceiving Simulated Ego-Motions in Virtual Reality - Comparing Large Screen Displays with HMDs. In: Proceedings of the SPIE, San Jose, CA, USA, vol. 5666, pp. 344–355 (2005)

Riecke, B.E., van Veen, H.A.H.C., Bülthoff, H.H.: Visual Homing Is Possible Without Landmarks: A Path Integration Study in Virtual Reality. Presence - Teleoperators and Virtual Environments 11(5), 443–473 (2002)

Riecke, B.E., von der Heyde, M., Bülthoff, H.H.: Visual cues can be sufficient for triggering automatic, reflex-like spatial updating. ACM Transactions on Applied Perception (TAP) 2(3), 183–215 (2005)

Riecke, B.E., Cunningham, D.W., Bülthoff, H.H.: Spatial updating in virtual reality: the sufficiency of visual information. Psychological Research 71(3), 298–313 (2007)

Ruddle, R.A., Payne, S.J., Jones, D.M.: Navigating large-scale virtual environments: What differences occur between helmet-mounted and desk-top displays? Presence - Teleoperators and Virtual Environments 8(2), 157–168 (1999)

Ruddle, R.A., Lessels, S.: For Efficient Navigational Search, Humans Require Full Physical Movement, but Not a Rich Visual Scene. Psychological Science 17(6), 460–465 (2006)

Ruddle, R.A., Lessels, S.: The benefits of using a walking interface to navigate virtual environments. ACM Trans. Comput.-Hum. Interact. 16(1), 1–18 (2009)

Tan, D.S., Gergle, D., Scupelli, P., Pausch, R.: Physically Large Displays Improve Performance on Spatial Tasks. ACM Transactions on Computer-Human Interaction (2005) (accepted)

Tan, D.S., Gergle, D., Scupelli, P., Pausch, R.: Physically large displays improve performance on spatial tasks. ACM Transactions on Computer-Human Interaction 13(1), 71–99 (2006)

Tan, D.S., Gergle, D., Scupelli, P.G., Pausch, R.: Physically large displays improve path integration in 3D virtual navigation tasks. In: CHI 2004: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 439–446. ACM Press, New York (2004)

Waller, D., Beall, A.C., Loomis, J.M.: Using Virtual Environments to Assess Directional Knowledge. Journal of Environmental Psychology 24(1), 105–116 (2004)

Williams, B., Narasimham, G., Westerman, C., Rieser, J., Bodenheimer, B.: Functional similarities in spatial representations between real and virtual environments. ACM Trans. Appl. Percept. 4(2), 1–22 (2007)

Witmer, B.G., Bailey, J.H., Knerr, B.W., Parsons, K.C.: Virtual spaces and real world places: Transfer of route knowledge. International Journal of Human-Computer Studies 45(4), 413–428 (1996)

Wraga, M., Creem-Regehr, S.H., Proffitt, D.R.: Spatial updating of virtual displays during self- and display rotation. Memory & Cognition 32(3), 399–415 (2004)

Eye Movements Reflect Reasoning with Mental Images but Not with Mental Models in Orientation Knowledge Tasks

Jan Frederik Sima, Maren Lindner, Holger Schultheis, and Thomas Barkowsky

SFB/TR 8 Spatial Cognition, Universität Bremen, Enrique-Schmidt-Str. 5,
28359 Bremen, Germany
{sima,mlindner,schulth,barkowsky}@sfbtr8.uni-bremen.de

Abstract. This paper presents results showing that eye movements reflect spatial relations in mental images but not in mental models during nearly similar reasoning tasks with directions. These results contribute to the distinction between mental models and mental images based on eye movements. This differentiation may be applied in the field of human-computer interaction and intelligent assistance systems. We conducted two experiments about reasoning with cardinal directions employing three-term series problems in the form of: "X is southwest of Z; Y is east of X; as seen from Z, where is Y?" The results replicate, to some extent, previous findings about preferred mental models. Additionally, the results indicate that these preferences are susceptible to details of the instructions of the experiment.

1 Introduction

This paper investigates the connection between reasoning with orientation knowledge and eye movements. We distinguish between two different mental representations that can underlie such a reasoning process, namely visual mental images (Kosslyn, 1994) and spatial mental models (Johnson-Laird, 1983; Tversky, 1993). In addition to previous eye tracking studies, which already found connections between eye movements and mental imagery, this study also explores the connection between eye movements and reasoning based on more abstract representations, that is, spatial mental models. To our knowledge, there are currently no studies that focused on eye movements in connection with spatial mental models. It is the aim of this study to investigate possible differences in eye movements during reasoning based on these different spatial representations. Spatial mental models are assumed to be abstract and amodal representations. In contrast, visual mental images are quasi-pictorial representations containing vivid details. Mental images can be described as resembling, at least to some degree, the experience of actually visually perceiving an object or a scene (Finke 1989). We hypothesize that there are less eye movements reflecting the content of a mental model than that of a mental image. Being able to distinguish these different representations during spatial reasoning tasks potentially provides new possibilities for the field of human computer interaction. For instance, if a computational assistance system is aware of the representation format the human reasoner employs, the system will be in a better position to provide appropriate assistance.

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 248–261, 2010.

Following up on the results of Bertel, Schultheis, and Barkowsky (2010), the conducted experiments also aim at exploring possible preferences in human reasoning about orientation relations. In this paper we compare two representation formats for deductive reasoning tasks with cardinal direction knowledge. The tasks are presented as three-term series problems (Johnson-Laird, 1972). Orientation knowledge can be considered a fundamental spatial knowledge type. Other fundamental spatial knowledge types are distance or topology. In this paper we focus on cardinal directions (e.g., north, northeast, east, etc.) as one form of orientation knowledge. Cardinal directions are qualitative relations given in terms of a global, geocentric reference system.

1.1 Reasoning with Orientation Knowledge in Under-Determined Tasks

Some reasoning tasks are not fully specified, that is, the given premises describe more than one spatial situation, thus, the problem is under-determined. In reasoning with orientation knowledge it is not possible to infer new orientation relations unambiguously when only orientation, but no distance relations between the entities are provided (Frank, 1996). A mental model, by definition, only represents one single instantiation of the described situation (Johnson-Laird, 1989). Several empirical studies in the field of spatial cognition showed that in the case of under-determined reasoning problems humans tend to construct only one mental model at a time to solve a problem rather than to construct all possible models (Knauff, Rauh, & Schlieder, 1995; Jahn, Johnson-Laird, & Knauff, 2005). These mental models, that are constructed primarily, have been termed *preferred mental models*. It is assumed that preferred mental models are easier to generate and faster to process and that they thus provide an efficient way to reason even with under-determined spatial information. Exploration of preferred mental models can accordingly make an important contribution to the understanding of human spatial reasoning.

Bertel et al. (2010) carried out an experiment to investigate preferred mental models in human deductive reasoning with cardinal directions. They had subjects work on three-term series problems containing two out of eight cardinal directions. An example of such a task would be: X is southwest of Z (1st premise), Y is east of X (2nd premise), as seen from Z, where is Y? (question). The given problems were restricted to those in which fictive lines between X and Z and between Y and X form a 45° (e.g., X is southwest of Z, Y is east of X) or a 90° angle (e.g., X is south of Z, Y is east of X). Given an eight-sector model of cardinal directions as the basis (see Fig. 1), four (for the 45° problems) or three (for the 90° problems) of the eight directions are correct answers; i.e., the reasoning problem is under-determined.

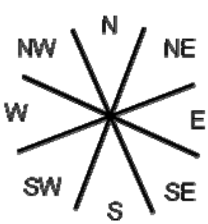


Fig. 1. Eight-sector model of cardinal directions

Figure 2 shows a taxonomy of possible mental models underlying these correct answers. Based on the subject's answer, the used mental model can be classified. If the distances between the entities are assumed to be equal, model ***b*** is constructed to solve the task. The answer would therefore be "southeast". If the distance between X and Z is longer than the distance between X and Y, model ***c*** is used and the according answer would be "south". Model ***a*** in which the distance between X and Z is considerably shorter than between X and Y results in the answer "east". On the contrary, if the distance between X and Z is considerably longer than the distance between X and Y, model ***d*** is used, which leads to the answer "southwest".

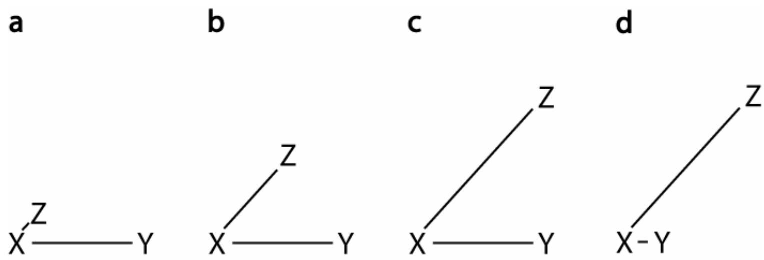


Fig. 2. Different possible mental models to solve the reasoning task "X is southwest of Z; Y is east of X; as seen from Z, where is Y?" The triangles of model ***b*** and ***c*** are considered prototypical and model ***c*** is the preferred model as the to be inferred direction is "south" compared to "southeast" in model ***b***.

Bertel et al. (2010) hypothesize that there are preferred mental models in reasoning with cardinal directions and that models will be preferred which a) form a prototypical triangle to constitute the problem situation and b) lead to one of the four main cardinal directions, i.e., north, east, south or west, as the inferred direction. Furthermore, property b) is weighted stronger than property a). Figure 2 gives an example of this preference. Thus, these assumptions lead to one hypothesized specific preferred model for each of the reasoning problems under investigation. Figure 3 illustrates two examples for hypothesized preferred mental models of 45° problems (1 and 2) and two for 90° problems (3 and 4).

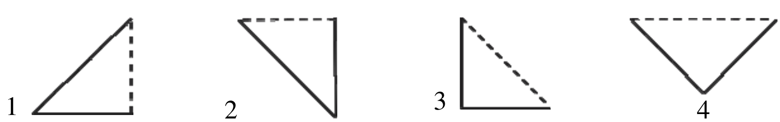


Fig. 3. Examples of hypothesized preferred mental models for 45° problems (1 and 2) and for 90° problems (3 and 4)

In the rest of this paper and in particular in the analysis we will use the following terms to distinguish between the possible models. The model that forms a prototypical triangle and leads to a main cardinal direction as the answer will be called "preferred model". The model that only forms a prototypical triangle will by called "prototypical model”. For 90° problems there are no prototypical models as those are always also preferred. The other models, which form distorted triangles, will be referred to as “distorted models”.

1.2 Mental Representations and Eye Movements

It has been argued that eye movements are linked to processes of attentional control (Shepherd, Findlay, & Hockey, 1986) and that they can be used as an indication of what the subject pays attention to and what he or she is concerned with in a particular task. Furthermore, there is a connection between visual information processing and eye movements as they are linked with visual brain areas by a projection path from extrastriate visual areas via the posterior part of the parietal cortex to regions associated with eye movements (Knauff, 2009). Several studies used eye tracking methodology to investigate mental images and showed a direct relation between eye movements and the processing of spatial relations in an imagined scene (e.g., Johansson, Holsanova, & Holmqvist, 2006). On the one hand, there are studies that show similarities between eye movements while subjects look at a particular picture and eye movements while these subjects imagine the previously seen picture (Brandt & Stark, 1997; Laeng & Teodorescu, 2001). On the other hand, there are experiments tracking the eye movements of subjects during the imagination of a previously unknown scene (Demarais & Cohen, 1998; Johansson et al., 2006; Ragni, Fangmeier, Bittner, & Konieczny, 2009; Spivey & Geng, 2001). In both cases there is clear correspondence between the eye movement and the mental image. There also are studies that show connections between eye movements and spatial relations in acoustically presented reasoning tasks (Demarais & Cohen, 1998; Ragni et al., 2009).

All these studies either presented the stimuli before the imagery phase or employed stimuli that are very easy to visualize and not abstract. For example, subjects had to imagine a colored fish (Laeng & Teodorescu, 2001), a forty story building with different actions happening on each floor (Spivey & Geng, 2001) or a scene rich of details and with many objects and spatial relations (Johansson et al., 2006). The reasoning tasks of Demarais and Cohen (1998) were also designed for easy visualization, e.g., by using household objects arranged in a shelf like “a jar of pickles is below a box of tea bags; the jar of pickles is above a can of coffee; where’s the can of coffee?”

It can be assumed that the way the experimental tasks are presented to a subject influences the mental representation used to solve these tasks. According to the definitions of mental images (Finke, 1989; Kosslyn, 1994), we state that if the stimuli are easy to visualize and contain a lot of vivid and visual details, subjects will employ visual mental images. In line with the definition of mental models (Johnson-Laird, 1983), we argue that if the stimuli are abstract and amodal, subjects will employ a non-visual but spatial representation, i.e., a spatial mental model. In order to grade stimuli as being more visual or more abstract, we refer to Schultheis, Bertel, Barkowsky, and Seifert (2007); they defined criteria to classify representations as being rather abstract and spatial or as being more vivid and visual. One criterion is for example exemplarity. That is, the more the stimulus is an example of a known situation (compared to a category or an abstract prototype) the more visual the used representation will be. Our assumptions are in line with the theory of deductive reasoning with mental models and visual mental images of Knauff (2009):

> "Visual brain areas are only involved if the problem information is easy to visualize and when this information must be processed and maintained in visual working memory. A regular reasoning process, however, does not involve visual images but more abstract spatial representations – spatial mental models – held in parietal cortices." (p. 111)

The discussed eye tracking studies mostly state that they investigate mental imagery and all of them used relatively easy-to-visualize stimuli. Additionally, they all reported eye movements according to the spatial relations of the stimuli. This shows that processing of mental images seemingly correlates with according eye movements. However, there seem to be no studies that report eye movements that reflect the spatial relations of spatial mental models. We assume that there is no systematic relation between eye movements and reasoning with spatial mental models. It is one aim of this contribution to test this assumption. Accordingly, we present two eye tracking experiments with similar reasoning tasks for which subjects are assumed to construct mental models or mental images, respectively.

2 Experiment 1 – Reasoning with Mental Models

The stimulus materials as well as the hypotheses regarding preferred mental models are as detailed in Bertel et al. (2010). It is the aim of experiment 1 to investigate the connection between eye movements and reasoning based on mental models. The reasoning problems used in this experiment are very abstract and not easy to visualize and it is assumed that subjects will thus use spatial mental models. It is further hypothesized that eye movements are not connected with reasoning based on spatial mental models, and that eye movements will therefore not occur significantly more often according to the directions given in the tasks compared to all other directions.

2.1 Participants and Materials

25 undergraduate students of the University of Bremen, 11 male and 14 female, volunteered to take part in the experiment for monetary compensation.

The three-term series problems used in the experiments contained the eight cardinal directions north, northwest, west, southwest, south, southeast, east, and northeast. Each trial is composed of two premises about two cardinal directions between three entities. The third direction relation is to be concluded. The entities are labeled with letters. The following is an example trial: "X is southwest of Z; Y is east of X; as seen from Z, where is Y?"[1].

For every problem trial the directions were chosen so that a line between X and Z and a line between Y and X would form either a 45° or a 90° angle. Due to the restriction of eight cardinal directions and problems with only 45° and 90° angles, there are 16 different problems. For each of the 16 problems it is possible to construct multiple different configurations and corresponding solutions, which all satisfy the premises. That is, the problems are under-determined. Using the example trial above, there are four possible different mental models, which are illustrated in Fig. 2.

[1] The language of the experiment was German. Tasks are translated into English here.

Depending on the order of the named entities, there are multiple possibilities to present each problem. The example above could also be given as a) “X is southwest of Z” and “Y is east of X” or as b) “Z is northeast of X” and “Y is east of X”, for example. As Bertel et al. (2010) found no indication that the order of the entities impacts the type of mental model that is constructed, the tasks of this study are only presented in one order, namely in the way described in b) above.

We employed a head-mounted SensoMotoric Instruments (SMI) iView X HED eye tracking system to record the subject’s eye movements.

2.2 Procedure

The participating subjects sat on a chair at a table facing a blank white wall at a distance of approximately 1m. Subjects had to place their hands on their legs under a table holding a computer mouse in the one hand and a small ball in the other one. This prevented the subjects from using their fingers to solve the tasks. The eye tracker was fixed on the subject’s head and calibrated. Subjects were told that the eye tracker's camera measures their pupil dilation in order to relate it to the difficulty of the tasks, i.e., they were not aware that their eye movements were recorded. All instructions of the experiment were projected on the white wall.

To make subjects familiar with the cardinal directions, the experiment started with a learning phase. Each task of the learning phase comprised an acoustically presented statement and an answer screen with a question. Each statement was triggered by the subject clicking her mouse and contained the direction relation between two entities; e.g., “K is northwest of U”. After 4 seconds the answer screen appeared, which depicted the reference entity “U” surrounded by the numbers 1 to 8 in a counterclockwise circular order together with the question “As seen from U, where is K?”. The eight numbers represented the eight cardinal directions (1 = north, 2 = northwest, 3 = west, … 8 = northeast). The participants answered by naming the number, which corresponded to the position of “K” and the experimenter typed the number into the computer. Depending on whether the answer was correct, “correct” or “wrong” together with the correct direction were projected. The learning phase ended as soon as each of the eight cardinal directions had been recognized correctly twice in a row.

The experiment consisted of 32 problem trials. These consisted of two instances for each of the 16 possible problems. The instances differed in the letters used. In addition, 4 pre trials were presented in the beginning in order to familiarize the subjects with the tasks. To prevent memory effects due to the identical order of the letters within in the problem trials, 12 filler trials with a different order were mixed in randomly. Thus, participants had to work on 48 trials in total, whereas only the 32 actual problem trials were used in the analysis. The trials were presented in a randomized order. The subjects used the mouse to trigger the acoustic presentation of the first premise of each trial. As soon as the subjects understood the statement, they clicked again for the presentation of the second premise. Similarly, they triggered the acoustic presentation of the question after having processed the second premise. When the subject felt she found the answer, she clicked the mouse again for the projected answer screen to appear. The answer screen and the answer procedure were the same as in the training phase, except that the question was not projected on the screen. To continue to the next trial, the subject had to click again. Subjects were allowed to take breaks between the trials.

The subject's eye movements were recorded during all tasks. When participants closed their eyes, the experimenter advised the subject to open her eyes again. The calibration, the learning phase and the test phase together typically lasted between 45 and 60 minutes.

2.3 Analysis of the Eye Tracking Data

In this section we will describe how the eye tracking data was analyzed to identify whether eye movements occurred along the given spatial relations between the entities for each trial. The same method of analysis was employed in both experiments. The raw eye tracking data collected by the iView X software was first converted using the IDF Event Detector to generate the list of fixations made by the subject. For this we had to set two parameters, maximum dispersion and maximum duration, which together define when a fixation is identified. Taking into account the distance between the subject and the white wall, we set these at 40 pts and 180 ms given a resolution of 752 x 480 of the tracked visual field. From the sequence and coordinates of the subject's fixations, the saccades were calculated automatically, as between two fixations at different locations, there must have been a saccade in order for the gaze to get from one point to the other.

Using the sequence of saccades, defined by the starting and ending coordinates, we classified each saccade into one of eight classes, which correspond to the eight cardinal directions, which were used in the trials: we uniformly mapped all possible angles of a saccade, which is basically a vector in a Cartesian plane and thus has an angle, to the cardinal directions. Each direction corresponds to a range of angles on a degree circle with each direction taking up (360°/8) = 45°. For example "north" corresponds to all angles in the range of 0° ± (45°/2) = 0° ± 22,5° = [337.5°; 22.5°]. It is to note that the eye movements classified in this way are relative eye movements, i.e., the absolute coordinates do not matter. This makes sense considering the fact that the subject moves her head during the trials and that also arbitrary eye movements occur in between. Given this classification of each saccade to one cardinal direction, we were able to investigate the relation between the given direction and observed eye movements during a trial.

We have so far only looked at the eye movements that occurred after the first premise of each problem trial was read. In particular, we analyzed only the eye movements that occurred in the second half of the time span between the first mention of the direction of the first premise and the time at which the subject clicks to indicate she has understood the premise and is ready for the second premise. We have discarded the first half of this time span, as we assume the subject needs to first process the presented information before she starts the construction of either a mental model or mental image.

As we expected eye movements to occur not only in the direction read, but also in the opposite one, we checked the set of saccades for both of these directions. Assuming a mental image of for example A being north of B, we should not only expect internal attention shifts from A to B but also from B to A during inspection as well as construction of the image. Thus, we always compare the absolute number of made saccades to the combined absolute number of saccades made along the given direction in the first premise and its opposite direction in the analysis of both experiments.

2.4 Results of Experiment 1

Accuracy

Despite the reported high difficulty of the task, the overall error rate was relatively low with an average error of 12.4%. The error rate ranged from 0% to 28.1%. Answers were counted as errors if they did not match any of the possible directions given the premises. Only the correct answers are relevant for the investigation of preferences, thus every incorrect answer was excluded from the analysis.

Eye movements

The proportion of eye movements along the given direction of the first premise and its opposite direction ranged from 14.29% to 34.88% with an average of 25.38% over all subjects. Assuming the directions given in the tasks do not have any influence on the eye movements, one would expect the eye movements being equally distributed over the eight cardinal directions, that is a proportion of 12.5% for each direction or 25% for the combination of a direction and its opposite direction. Given our average result of 25.38%, there are not substantially more eye movements according to the named directions and its opposites. Binomial tests showed that only 1 subject out of 25 (see Fig. 4) made significantly[2] more eye movements along the directions of the first premise. None of the other subjects moved their eyes significantly more often along the given directions. This is what one would expect, if – as hypothesized – eye movements do not reflect the reasoning process.

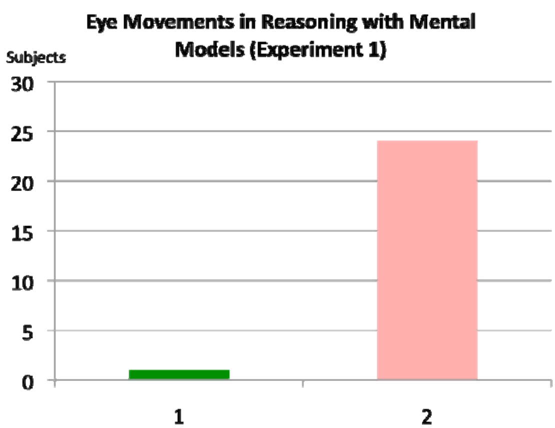


Fig. 4. The number of subjects that made significantly more eye movements along the directions given in the first premises and their opposites (1) and those that did not (2) during experiment 1

Preferred mental models

Every answer was classified as either being consistent or not consistent with the hypothesized preferred mental models (see section 1.2). For each subject the percentages

[2] If not stated otherwise, the level of significance is p=0.05.

of how often the subject constructed the hypothesized preferred or other mental models were computed. These percentages served as the basis for subsequent analyses. Since percentages are often not distributed normally, non-parametric test were employed whenever this was necessary and possible.

If there were no preferences, the probability of a subject choosing any possible correct answer would be uniform. For each 45° problem, the hypothesized preferred solutions should then be constructed on 25% of all trials, as there are four correct answers. For the 90° problems, there are three different correct answers, including the hypothesized preferred one. That is, each solution should be chosen on 33% of the trials, given there are no preferences.

Regarding the 90° problems, the hypothesized preferred models were chosen on 88% of all trials which indicates a strong preference for these models (see Fig. 5). Indeed, the preference is much stronger than one would expect by chance ($t_{(24)} = 17,551$; $p < 0.001$). We can conclude, that the results confirm the hypothesis regarding preferred mental models for the 90° problems.

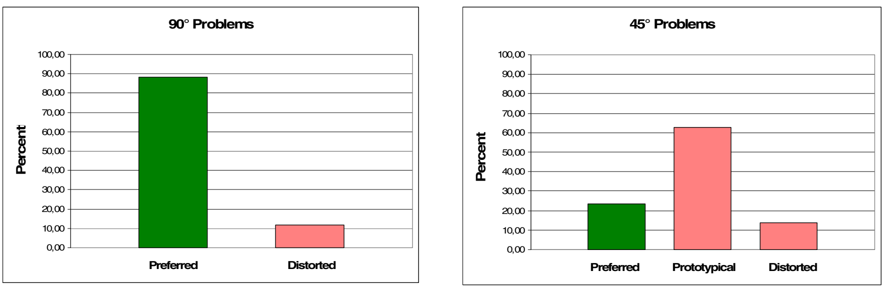


Fig. 5. Average percentages of trials on which subjects constructed the hypothesized preferred mental models compared to those that constructed other correct models for the 90° and the 45° problems of experiment 1

Regarding 45° problems, the preferred, prototypical, and distorted models were chosen on 23%, 63%, and 14% of the trials, respectively. Statistical analyses revealed that subjects exhibited no preference for the hypothesized preferred models ($t_{(24)} = -1.021$; $p > 0.3$), a significant preference for prototypical models ($t_{(24)} = 5.352$; $p < 0.001$), and a significant dispreference for distorted models ($t_{(24)} = -9.995$; $p < 0.001$). The strong preference for prototypical models is further corroborated by pairwise comparisons using the Wilcoxon test: Prototypical models are chosen more frequently than both preferred ($z = -2.145$; $p < 0.05$) and distorted ($z = -3.704$; $p < 0.001$) models while frequencies for preferred and distorted models do not differ significantly from each other ($z = -0.392$; $p > 0.65$). Concluding, the results do not confirm the hypothesis regarding preferred mental models for the 45° problems. Thus, results of Bertel et al. (2010) regarding the 45° problems could not be replicated. This may be due to the different experimental settings. Subjects in Bertel et al. (2010) were presented the premises and questions on a monitor whereas in our experiments they were facing a blank white wall and had to keep their hands below the table while the premises were presented acoustically.

3 Experiment 2 – Reasoning with Mental Images

In this experiment we induced the use of mental imagery for reasoning about cardinal directions. We assume that by using stimuli, that are easier to visualize, subjects will employ visual representations to solve the tasks, i.e., visual mental images. We hypothesize, that during a reasoning process based on mental images, significantly more eye movements will occur along the cardinal directions given in the tasks than along the other directions. Besides the minor modification of the instructions, this experiment is identical to the first one and reused the reasoning problems as well as the hypotheses regarding preferred mental models of experiment 1.

3.1 Participants and Materials

23 undergraduate students of the University of Bremen, 12 male and 11 female, participated in the study for monetary compensation. The stimulus material used in this experiment is the same as in experiment 1 and is described in section 2.1.

3.2 Procedure

The procedure of this experiment resembles that of experiment 1 (see section 2.2). The only difference is a slight change in the instructions given to the participants. In contrast to experiment 1, the subjects were told that the entities in the experimental tasks are meant to be cities on a map. Furthermore, they were told that each city is marked on the map as a little red square with the according letter written next to it labeling the city. Subjects were instructed to imagine the given constellation of the three cities on the imaginary map to solve the reasoning problem. These instructions were used for the learning phase as well as for the test phase. By describing the stimuli in a more visual way than in the first experiment, we wanted to induce the use of mental imagery. The acoustic presentation of the stimuli during the trials remained unchanged.

3.3 Results of Experiment 2

Accuracy
In this experiment, the error rate ranged from 0% to 28.1% with an average of 9.5%.

Eye movements
We proceeded as in the previous experiment, described in section 2.3 and 2.4. We related the proportion of eye movements that occurred according to the cardinal directions given in the premises and their corresponding opposite directions, which was 30.46% (with a range from 16.46% to 50.40%) over all subjects, to the level of chance, namely 25%. Binomial tests showed that 10 participants made significantly more eye movements along the expected directions than along the other directions, whereas 13 subjects did not (see Fig. 6). This number of 10 participants is significantly more than expected, given the $p=0.05$ probability of error when testing the individual subject. These results confirm our hypothesis about the correlation between eye movements and the spatial relations in reasoning with visual mental images.

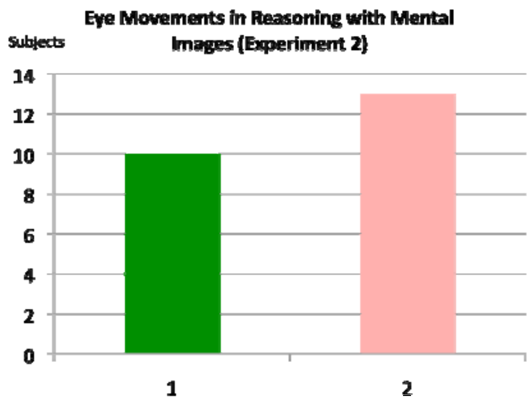


Fig. 6. Comparison of subjects that made significantly more eye movements along the directions in the premises (1) and those that did not (2) during experiment 2

Preferred mental models

On average, 93% hypothesized preferred mental models were constructed for 90° problems (see Fig. 7). As in experiment 1, the preference is much stronger than one would expect if participants had no preferences ($t_{(22)}$ = 29,846; $p < 0.001$).

Regarding the 45° problems, the preferred, prototypical, and distorted models were chosen on 48%, 46%, and 6% of the trials, respectively. Statistical analyses revealed that all these percentages differed significantly from percentages expected if people had no preferences ($t_{(22)} = 2.683$; $p< 0.05$; $t_{(22)} = 2.512$; $p< 0.05$; $t_{(22)} = -25.36$; $p< 0.001$ for preferred, prototypical, and distorted models, respectively). Thus, participants exhibited marked preferences for both preferred and prototypical models. Pairwise comparisons using the Wilcoxon test indicate that there is no difference in preference for preferred and prototypical models ($z = -0.305$; $p > 0.75$), while both of these are chosen significantly more frequently than distorted models ($z = -3.569$; $p < 0.001$ and $z = -3.262$; $p < 0.001$, respectively). Consequently, the results of Bertel

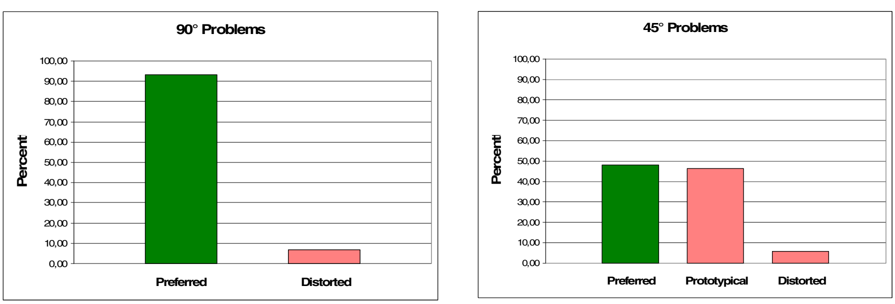


Fig. 7. Average percentages of trials on which subjects constructed the hypothesized preferred mental models compared to those that constructed other correct models for the 90° and the 45° problems of experiment 2

et al. (2010) were largely replicated, showing a reliable preference for the hypothesized preferred mental models.

4 Comparing Eye Movements and Preferred Mental Models in Experiment 1 and 2

One main goal of this investigation was to explore the occurrence of eye movements in reasoning with cardinal directions based on mental models and mental images. We hypothesized that eye movements according to the spatial relations of the tasks will occur when employing visual mental images (as in experiment 2) and that this will not be the case when employing spatial mental models (as in experiment 1). Results show that in the mental image condition 30.56% of the eye movements occurred according to the directions and respective opposite directions of the tasks (see section 3.3). This is well above the level of change of 25%. In contrast, the proportion of eye movements in line with the directions of the premise when using spatial mental models was considerably lower with only 25.38%, which equals the level of chance (see section 2.4). This comparison already indicates an influence of the representation format on eye movements during the tasks.

In addition, as described in section 3.3, there are significantly more subjects (10 out of 23) in the mental imagery experiment that showed significantly more eye movements along the directions in the task. In contrast only 1 subject out of 25 (see section 2.4) showed significantly more eye movements along the given directions during the mental model experiment. A χ^2 test showed that this difference between the two experiments is significant ($p < 0.01$). This means that there is a clear influence of the used mental representation on eye movements during spatial reasoning tasks.

One could think of an alternative explanation for the different eye movement patterns in the two experiments. The change in instruction could have led the subjects to use a different perspective for the two conditions. Specifically, subjects could have used an egocentric perspective for solving the tasks of the first experiment and a survey perspective for the second one, for which the image of a map is suggested in the instructions.[3] We do, however, assume no change in perspective between the two conditions, as subjects are trained to apply a survey perspective in both cases as described in section 2.2.

Regarding the preferred mental models, both experimental conditions yield similar results for the 90° problems, but there is a notable difference for the 45° problems. Our hypothesized preference for those problems seems to only apply when mental imagery is used to solve the tasks. In the mental model condition the prototypical model was actually constructed significantly more often than the other models. This indicates that the preferred mental model may depend on the underlying mental representation that is used.

5 Conclusions

This contribution reports two experiments that investigated human spatial reasoning about cardinal directions. Of particular interest was the influence the representation

[3] We thank one of the reviewers for pointing this out.

format employed for reasoning has on the eye movements during reasoning. The experimental results have a number of important implications. First, the existence of preferences in both experiments (a) further corroborates the idea that preferences are robust and pervasive phenomena in human spatial reasoning and (b) suggests that preferences also occur when the employed representations are more image-like (i.e., preferred mental images). Second, the fact that the type of preferences observed in the present studies partly differ from previously observed preferences in reasoning about cardinal directions suggests an unanticipated susceptibility of the type of preferences to procedural details. Third, when employing an abstract representation format such as spatial mental models, eye movements do not reflect the spatial layout of the represented situation. This is in contrast to eye movements observed in the scope of employing more visual representations: when more visual representations such as mental images are used, eye movements often reflect the spatial relations that are represented. Fourth, the differing impact that visual and abstract representations have on eye movements opens up the possibility to utilize observed eye movements as evidence for the type of representation currently employed by a person. This could be of advantage in building or improving computer-based assistance systems that support a human reasoner in solving spatial problems such as, for instance in spatial planning or in architectural design. If the assistance system receives evidence about the format in which the reasoner currently represents the task-relevant knowledge, the system will be able to better predict the cognitive state of the reasoner (e.g., based on an available cognitive model) and, thus, will be able to provide better assistance.

Acknowledgements

This paper presents work done in the project R1-[ImageSpace] of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition. Funding by the German Research Foundation (DFG) is gratefully acknowledged. We also thank Sven Bertel for his suggestions regarding experimental design.

References

Bertel, S., Schultheis, H., Barkowsky, T.: Modeling mental spatial reasoning about cardinal directions. Manuscript submitted for publication (2010)

Brandt, S.A., Stark, L.W.: Spontaneous eye movements during visual imagery reflect the content of the visual scene. Journal of Cognitive Neuroscience 9(1), 27–38 (1997)

Brockmole, J.R., Wang, R.F.: Switching between environmental representations in memory. Cognition 83, 295–316 (2002)

Demarais, A.M., Cohen, B.H.: Evidence for image-scannig eye movements during transitive inference. Biological Psychology 49, 229–247 (1998)

Finke, R.: Principles of mental imagery. MIT-Press, Cambridge (1989)

Frank, A.U.: Qualitative spatial reasoning: Cardinal directions as an example. International Journal of Geographical Information Science 10(3), 269–290 (1996)

Jahn, G., Johnson-Laird, P.N., Knauff, M.: Reasoning about consistency with spatial mental models: Hidden and obvious indeterminacy in spatial descriptions. In: Freksa, C., Knauff, M., Krieg-Brückner, B., Nebel, B., Barkowsky, T. (eds.) Spatial Cognition IV. LNCS (LNAI), vol. 3343, pp. 165–180. Springer, Heidelberg (2005)

Johansson, R., Holsanova, J., Holmqvist, K.: Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. Cognitive Science 30, 1053–1079 (2006)

Johnson-Laird, P.N.: The three-term series problem. Cognition 1, 57–82 (1972)

Johnson-Laird, P.N.: Mental models. Harvard University Press, Cambridge (1983)

Johnson-Laird, P.N.: Mental models. In: Posner, M.I. (ed.) Foundations of cognitive science, pp. 469–499. MIT-Press, Cambridge (1989)

Knauff, M.: A neuro-cognitive theory of deductive relational reasoning with mental models and visual images. Spatial Cognition & Computation 9, 109–137 (2009)

Knauff, M., Rauh, R., Schlieder, C.: Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. In: Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, pp. 200–205. Erlbaum, Mahwah (1995)

Kosslyn, S.M.: Image and Brain: The Resolution of the Imagery Debate. Harvard University Press, Cambridge (1994)

Kosslyn, M.S., Alpert, M.N., Thompson, L.W., Maljkovic, V., Weise, B.S., Chabris, F., et al.: Visual mental imagery activates topographically organized visual cortex: PET investigations. Journal of Cognitive Neuroscience 5, 263–287 (1993)

Kosslyn, M.S., Thompson, L.W.: When is early visual cortex activated during visual mental imagery? Psychological Bulletin 129(5), 723–746 (2003)

Laeng, B., Teodorescu, D.-S.: Eye scanpaths during visual imagery reenact those of perception of the same visual scene. Cognitive Science 26, 207–231 (2001)

Ragni, M., Fangmeier, T., Bittner, A., Konieczny, L.: Incremental model construction: Eye-movements reflect mental representations and operations – even if there is nothing to look at. In: Proceedings of the 31st Annual Conference of the Cognitive Science Society, pp. 3046–3051. Cognitive Science Society, Austin (2009)

Schultheis, H., Bertel, S., Barkowsky, T., Seifert, I.: The spatial and the visual in mental spatial reasoning: An ill-posed distinction. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, R.D. (eds.) Spatial Cognition 2007. LNCS (LNAI), vol. 4387, pp. 191–209. Springer, Heidelberg (2007)

Shepherd, M., Findlay, J.M., Hockey, R.J.: The relationship between eye movements and spatial attention. Quarterly Journal of Experimental Psychology 38A, 475–491 (1986)

Sloman, A.: Afterthoughts on analogical representation. In: Proc. Theoretical Issues in Natural Language Processing (TINLAP-1), Cambridge, MA, pp. 164–168 (1975)

Spivey, M.J., Geng, J.J.: Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. Psychological Research 65, 235–241 (2001)

Tversky, B.: Cognitive maps, cognitive collages, and spatial mental models. In: Frank, A.U., Campari, I. (eds.) COSIT 1993. LNCS, vol. 716, pp. 14–24. Springer, Heidelberg (1993)

An Eye-Tracking Study of Integrative Spatial Cognition over Diagrammatic Representations

Atsushi Shimojima[1] and Yasuhiro Katagiri[2]

[1] Doshisha University, 1-3 Tatara-Miyakodani, Kyotanabe, Kyoto, Japan
[2] Future University Hakodate, 116-2 Kamedanakano, Hakodate, Hokkaido, Japan

Abstract. Spatial representations, such as maps, charts, and graphs, convey different levels of information, depending on how their elements are grouped into different units of objects. Therefore, how people set boundaries to graphical objects to be interpreted and how they maintain the object boundaries during the given task are two important problems in understanding the way people utilize spatial representations for problem-solving. Table comprehension process was experimentally investigated in terms of eye gaze control behaviors when people were required to read off information distributed over large-scale objects, e.g., a row or a column, of the given table. Evidence was found that a large-scale object can be bounded by a single attentional shift to it, and that they can be retained as coherent objects for subsequent reference. These findings suggest the existence of a higher-order information processing in the comprehension of a spatial representation, based on rather intricate processes of attention management.

Keywords: spatial representation, object-based attention, visual index, eye-tracking, embodied cognition, situated cognition.

1 Introduction

Reasoning over external spatial representations, such as maps, graphs and charts, plays a significant role in human spatial cognition. One of the strong characteristics of human higher cognition lies in its capacity in the application of spatial representations that is extended to operate in non-spatial domains.

This aspect of human cognition, the acquisition and the utilization of knowledge in abstract spatial environments, has mostly been studied in the area of diagrammatic reasoning. Even though actual human problem solving with diagrammatic representations involves intricate coordination of both acquisition and utilization aspects of spatial information, emphasis has been placed mostly on the utilization aspect. Extensive studies have been devoted to the understanding of the semantic matching between spatial representations and non-spatial represented domains [2,30], how spatial constraints are effectively utilized in reasoning about non-spatial facts [25], and the characterization and development of human spatial skills [9]. In these studies, it has tacitly been assumed that all the relevant internal spatial information is acquired from the external spatial representations before cognitive reasoning processes take place.

In contrast to this static and detached picture in diagrammatic reasoning of acquisition and utilization of spatial information, studies in spatial cognition in real spatial

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 262–278, 2010.

environments, either in human way-finding or in robot navigation, emphasizes more active and interactive nature of underlying cognitive processes. Cognitive agents are embedded in spatial environments, and the acquisition and the utilization of spatial information is tightly intertwined. Computational models have been developed to describe and simulate cognitive processes that try to capture this situated and embodied nature of human spatial cognition [31].

We attempt, in this paper, to combine these two research traditions, and to establish more situated and embodied pictures of cognitive processes in human problem solving with spatial representations. Our approach is defined by two leading questions and two corresponding hypotheses, so we will begin with formulating them in detail (section 2). We will then describe an eye-tracking experiment on people engaged in table-reading tasks (sections 3 and 4). The tasks are designed so that the levels of information to be extracted from the given spatial representation may be apparent. The eye-tracking data offer an integrative view of the use of spatial representation, where attentional mechanism affects higher-level cognitive processes in a profound manner (section 5).

2 Problem

Researchers and designers have noted that information graphics can express "higher-level" information as well as "lower-level" information [3,4,32,10]. The former roughly indicates more abstract information carried by overall patterns formed by multiple graphical elements, while the latter more concrete information carried by individual graphical elements.

For example, the location of individual dots in Figure 1 indicate the existence of individual data points with specific values. In addition to this lower-level information, the scatter plot expresses higher level-information by "the shape and the density of the cloud" formed by these dots [12]. While the lower-level information is concerned with the values taken by individual cases in the data, the higher-level information is concerned with the overall distribution of the data, such as the strength of correlation between the two variables.

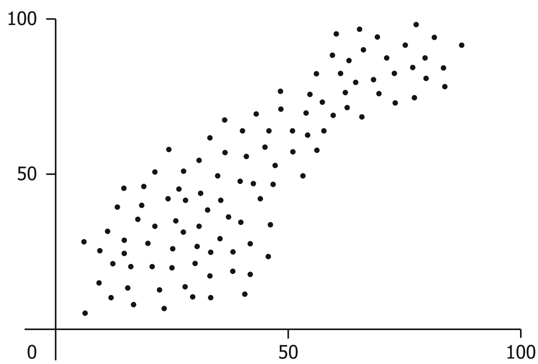


Fig. 1. A typical scatter plot

The distinction of higher-level and lower-level information is quite general over different kinds of spatial representations. Just as clouds of dots in a scatter plot can have informative shapes, line segments in a line graph can form an informative slope or curve by connecting individual points on a plane that carry lower-level information. Likewise, when the bars in a bar chart has the shape of "descending staircase" [18], this can mean that the price of a product steadily declined during the period.

Data maps have been also cited as carriers of multiple-level information. For example, Lowe [15] discussed "secondary structure", where adjacent isobars on a meteorological map together indicate a global trend of the area's barometric situation. Gilhooly et al. [8] found the use of "specialist schemata" in geographers' reading of contour maps, where visual patterns formed by several contour lines indicate some global structures in the area, such as valleys and interlocking spurs. Ratwani et al. [21] distinguished "specific information" and "integrating information" that can be extracted from choropleth graphs, the kind of data maps using color and shading of regions to represent magnitude.

Node-edge graphs and even tables support higher-level information. Olivier [17] discussed the case of tree diagrams, where an extended path formed by consecutive edges indicates the presence of a descent or chain in the represented relational structure. In London's tube map, the concentration of edges touching a node indicates the presence of a "hub" station [26]. Many tables are designed to allow the viewer to do "column-wise" or "row-wise" readings, in addition to basic "cell-wise" readings [26].

As the notion of "higher-level information" is applicable to such a wide variety of cases, one may suspect that it might be without content. Shimojima [26] investigated how certain spatial representations come to carry higher-level information, and identified a general pattern in which additional semantic rules are logically derived from basic semantic conventions in a spatial representation system. Thus, the level difference of information expressed by spatial representations is not just the matter of our subjective judgment, but susceptible to exact semantic characterization.

Nevertheless, whether a reader can appreciate different levels of information expressed in the given spatial representation is the matter of exact cognitive operations involved in the comprehension process. Specifically, which level of information people extract should be profoundly affected by the following factors:

Question (1) how people set boundaries to graphical objects to be interpreted.
Question (2) (when the reading task is complex enough) how they maintain object boundaries during the task.

For example, in order to extract specific values of individual data points from a scatter plot in Figure 1, one need only interpret the locations of individual dots, whereas one need take the entire cloud of dots as a coherent object in order to evaluate the overall trend of the data. In addition, when the task is to compare the strength of correlation in two data sets (say, the data in the *x*-range 0–50 and those in the x-range 50–100 in Figure 1), one need somehow maintain the boundaries of more than one higher-level graphical objects.

The purpose of the research presented in this paper is to begin investigations into the above two questions, by examining the following two fundamental hypotheses:

Hypothesis (1). A large-scale object in a spatial representation can be bounded by a single shift of attention to it, not necessarily through the integration of smaller component objects separately attended to.

Hypothesis (2). Large-scale objects, once bounded, can be retained as coherent objects, and do not have to be reintegrated from component objects when they are accessed subsequently.

The first hypothesis is motivated by the idea of "bounded activation," or "coloring," proposed by Ullman [31]. Ullman developed a theory of "visual routines," namely, sequences of elemental operations applied to particular locations in the visual scene to analyze visual features and spatial relations holding there. Bounded activation is one of the postulated elemental operations, whose function is to define coherent units of regions in the unarticulated visual scene so that further operations can be applied selectively to the activated regions. Roelfsema and his colleagues made this idea more exact by proposing computational and neurological models of the operation [23,22], and provided neuro-physiological evidence to its functioning in macaque monkey [24,14]. "Object-based attention" actively investigated by Duncan [7], Kramer and Jacobson [13] and Driver and Baylis [6] largely overlaps with the operation of bounded activation. Strong empirical evidence for the operation has been accumulated in this tradition too (e.g., O'Craven et al. [16]). Our hypothesis states that the operation of coloring, or object-based attention, can be applied to large-scale objects in spatial representation in order to extract task-relevant higher-level information from them.

The second hypothesis is motivated by the idea of "marking" proposed, again, by Ullman [31]. According to Ullman, some visual tasks require the application of elemental visual operations to multiple locations of the scene. Some tasks further require one to combine information obtained at different locations. The operation of marking is supposed to meet this demand by maintaining the record of the locations already visited and of summary information associated with these locations. Kahneman et al. [11] generalized this concept into the concept of "object files," by which we keep track of objects as coherent entities despite changes of their locations and visual features. Although it is still an open question how this operation is implemented computationally and neurologically, recent studies of visual indexing operations[19,20,1,29] seem to provide good evidence for the existence of cognitive mechanism with the object-tracking functionality.

Our second hypothesis states that large-scale objects can be tracked in this manner, with their locations and summary information retained in working memory for easy subsequent access. Indeed, Ratwani et al. [21] emphasized that graph comprehension in realistic setting involves what they called "cognitive integration," where multiple large-scale objects obtained from spatial representations are compared to each other for the extraction of higher-level information such as general trends of data. Such comparisons would require the retainment of large-scale objects after the initial bounding operations, and thus give an additional motivation to our hypothesis.

3 Methods

As an initial test of these hypotheses, an eye-tracking study was conducted on a group of participants who were engaged in table comprehension tasks.

3.1 Participants

A total of 46 students, 14 males and 32 females, participated in the experiment. Average age of participants were 20.4. There were 4 participants who failed to follow the instructions and their data were excluded from the analysis of response time and accuracy. The eye movements of additional 10 participants were not reliably extracted due to errors in the eye-tracking hardware, and their data were not included in the subsequent analysis of saccadic eye movements.

3.2 Materials

The tables presented to participants as visual stimuli were divided into two large groups. One group, called *black-and-white tables*, use black and white squares as their main symbols, whereas the other group, called *T-and-F tables*, use the roman letters "T" and "F" as their main symbols. Both kinds of tables express a membership relation from the people 1–5 to the organizations A–E. In a black-and-white table, the appearance of a black square means membership, whereas the appearance of a white square means non-membership. In a T-and-F table, the letter "T" indicates membership and the letter "F" indicates non-membership. Thus, the black-and-white table in Figure 2 lets us read off the information that person 3 is a member of organization C for example, whereas the T-and-F table in Figure 2 lets us read off the opposite information.

Each group of tables are further divided into 9 different kinds, depending on their horizontal and vertical spacing. Figure 3 shows the nine spacing patterns of our stimulus tables. "Neutral" spacing was approximately .7 deg viewing angle, while "sparse" spacing was twice as large (approximately 1.4 deg) and "dense" was one fifth as large (approximately .1 deg). Each symbol (square or letter) on a stimulus table was .6 deg wide and long, so the length of a column or row of a stimulus table was approximately 4.0 degrees under dense spacing, 6.3 degrees under neutral spacing, and 9.0 degrees under sparse spacing.

	A	B	C	D	E
1	■	■	□	□	□
2	□	□	□	■	■
3	□	□	■	■	□
4	■	□	□	□	■
5	□	■	□	□	□

	A	B	C	D	E
1	F	T	T	F	F
2	T	T	T	T	T
3	F	T	F	T	F
4	F	T	F	F	T
5	F	F	T	T	F

Fig. 2. A black-and-white table (left) and a T-and-F table (right) used as stimuli in the experiment

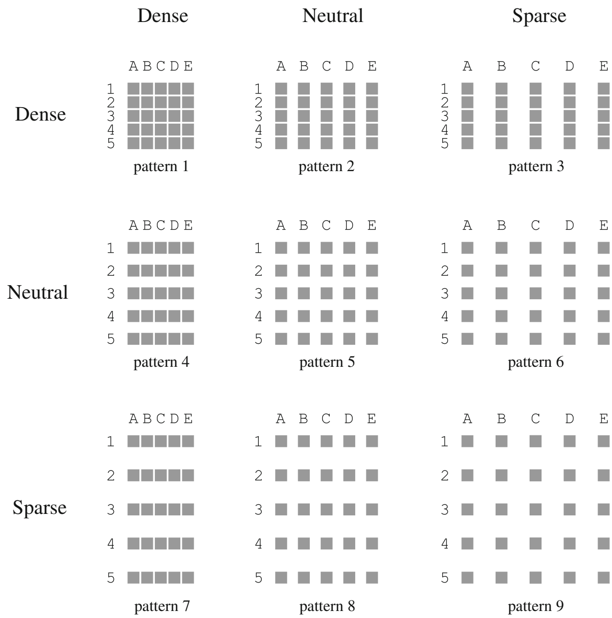


Fig. 3. Nine spacing patterns of the stimulus tables. Patterns 2, 3, and 6 have column-wise grouping of symbols, while patterns 4, 7, and 8 have row-wise grouping. Patterns 1, 5, and 9 have neutral grouping.

Stimulus tables of spacing type 2, 3, or 6 shown in Figure 3 have horizontal spacing larger than vertical spacing, making columns good perceptual groups. We say that such tables have *column-wise grouping*. Then, tables of spacing type 4, 7, or 8 are said to have *row-wise grouping*, and tables of spacing type 1, 5, or 9 are said to have *neutral grouping*.

3.3 Task

On the basis of these tables, two kinds of reading tasks were imposed on the participants. In the *row-wise task*, the participants were asked to read the given table to assess the number of people who belong to exactly two organizations. Suppose the black-and-white table in Figure 2 is shown with an instruction to solve this problem. How would one go about? Well, one should read the rows of the table one by one, making judgment whether it contains two black squares. Most probably, one would begin with the uppermost row representing person 1 and go down. If one goes right, one would reach the conclusion that there are four rows (representing persons 1 , 2, 3, and 4) containing exactly two black squares. One would then answer "four."

Thus, the rows of the table are the primary objects to be scanned and interpreted in this task, and we express this fact by saying that this task has the *horizontal axis*. In terms of our hypotheses, rows are large-scale objects to be bounded in this task, and a judgment whether the row contains exactly two black squares is the task-relevant information to be attached to each row. We want to examine whether rows can be bounded by a single shift of attention in this kind of tasks, and whether they can be stored as coherent objects for subsequent reference.

In contrast, the *column-wise task* asked the participants to assess the number of organizations to which exactly two people belong. Try to solve this problem on the basis of the black-and-white table in Figure 2. This time, the primary objects to be scanned and interpreted are columns of the table, and whether the column contains exactly two black squares is the task-relevant information to be attached to each column. Thus, this task is said to have the *vertical axis*. Well, there are four columns (representing organizations A, B, D, and E) containing exactly two black squares, and the correct answer is "four." The reader should have reached the same conclusion by scanning columns of the table one by one, perhaps starting with the leftmost column.

3.4 Procedure

Initial setup. After they were informed of the experiment and signed the consent form, participants were instructed to sit in front of a 19-inch display where the stimulus tables were to be presented. Their foreheads and chins were fixated, and a standard 9-point eye tracker calibration was conducted. The distance of eyes and the center of the display was approximately 50 cm.

Presentation steps. Task instruction was displayed in the beginning of each block of problems, followed by three exercise problems. Participants were asked to press, as quickly and as accurately as possible, a button with the number representing the answer they reached. They were instructed to press buttons without averting their eyes from the display.

A total of 36 different types of problems were prepared, varying in the task axis (vertical or horizontal), the table symbol (black-and-white or T-and-F), and the spacing of the table (see Figure 3). Each type of problems had 3 instances, totaling up to 108 problems to be presented to a participant. To ease the difficulty in distinguishing rather complex problem types, we presented problems in the sequence of 4 blocks, each consisting of the same type of problems defined by the task axis and the table symbol. The order of problems within a block was randomized.

Measurements. Eye movement patterns of participants while they were reading tables were recorded with an eye tracker NAC EMR-AT VOXER, which has 60Hz temporal precision. Fixation points were identified from eye samples with the centroid-based dispersion threshold method [5]. The threshold for maximum dispersion was set to 1.67 degrees in radius, while the threshold for minimum fixation duration was set to 100 milliseconds. Saccades were then identified as connecting two fixations. Percentage of correct answers and response times were recorded with SuperLab.

4 Predictions

Our first hypothesis states that a large-scale object in a spatial representation can be bounded by a single shift of attention to it. Driver and his colleagues conducted a series of experiments showing that object-based attention is directed to parts of the visual scene making good perceptual units defined by Gestalt principles (e.g., [6]). Thus, we expect that the hypothesized bounding of a large-scale object would happen when it is a primary object to be interpreted in the given task and it makes a good perceptual unit. In our experimental setting, columns in column-wise tables under column-wise tasks satisfy this condition. Rows in row-wise tables under row-wise tasks do so too. Thus, we predict that scanning eye movements along such columns or rows are minimal or even absent, since single attention shifts should be sufficient for setting their boundaries.

To operationalize the key concepts, we define a *horizontal saccade* as a saccadic eye movement with a larger displacement in the horizontal direction than in the vertical direction. It is considered to be an eye movement along the rows of the table. Correspondingly, a *vertical saccade* is a saccadic eye movement with a larger displacement in the vertical direction than in the horizontal direction. It is considered to be an eye movement along the columns of the table. We also call a saccade moving along the axis of the given task a *coaxial saccade*. Thus, vertical saccades occurring in a column-wise task and horizontal saccades occurring in a row-wise task are coaxial saccades in our sense. Our first prediction is then the following:

Prediction (1). Coaxial saccades are fewer when the table grouping matches with the task axis than when they mismatch, and the number of coaxial saccades in the matching condition approaches zero.

Our second hypothesis states that large-scale objects, once bounded, can be retained as coherent objects. This implies that large-scale objects can be revisited, when the need occurs to check the task-relevant information stored with it. When translated into eye movement patterns, *backward saccades* are predicted to occur from columns to columns in column-wise tasks, and from rows to rows in row-wise tasks. Such backward saccades, however, can occur only when columns or rows are retained as coherent objects. Thus, the frequency of backward saccades should depend on the table grouping. Specifically, if the table grouping matches with the task axis, and thus the columns or rows to be retained make good perceptual units, then they can be in fact retained as coherent objects and backward saccades can occur to them. In contrast, if the table grouping does not match the task axis, and thus the columns or rows to be retained fail to be good perceptual units, then some of the columns or rows are not retained as coherent objects in the first place. Lacking target coherent objects, the frequency of backward saccades will fall under such conditions.

What counts as a backward saccade depends on the axis of the task and the spacing of rows or columns in the given table. In the case of vertical tasks, a backward saccade is a leftward saccade whose horizontal displacement equals or exceeds the distance between the vertical axes of adjacent columns in the given table. Since our stimulus tables have different horizontal spacings between columns, the criterion of backward saccades depends on which spacing the given table has. A backward saccade in horizontal tasks is defined analogously, as an upward saccade whose vertical displacement equals or exceeds the distance between the horizontal axes of adjacent rows in the given table. With this definition at hand, our second prediction can be stated in the following way:

Prediction (2). Backward saccades will occur under all conditions, and their frequency will be higher when the table grouping matches with the task axis than when they mismatch.

5 Results

5.1 Coaxial Saccades

Table 1 shows the average number of coaxial saccades in column-wise tasks, relative to the table symbol and the table grouping. Table 2 does the same for coaxial saccades in row-wise tasks.

The main effect of the table grouping on the frequency of coaxial saccades was very strong both in column-wise tasks ($F(2, 62) = 182.3, p < .0001$) and in row-wise tasks ($F(2, 62) = 128.6, p < .0001$). Pairwise comparisons show that coaxial saccades in row-wise tasks were the least with row-wise tables, more with neutral tables, and the most with column-wise tables; the frequency of coaxial saccades in column-wise tasks had the opposite tendency. The main effect of the table symbol was also found, both for row-wise tasks ($F(1, 31) = 65.0, p < .0001$) and for column-wise tasks ($F(1, 31) = 192.7, p < .0001$). Strong interactions of the table symbol and the table grouping were also found for column-wise tasks ($F(2, 62) = 125.0, p < .0001$) and for row-wise tasks ($F(2, 62) = 38.2, p < .0001$).

5.2 Backward Saccades

Table 3 shows the average number of backward saccades in column-wise tasks, relative to the table symbol and the table grouping. Table 4 does the same for backward saccades in row-wise tasks.

Strong main effect of the table grouping was found on the frequency of backward saccades both for column-wise tasks ($F(2, 62) = 49.0, p < .0001$) and for row-wise tasks ($F(2, 62) = 10.3, p < .0001$). Overall, more backward saccades were made when the table grouping matched with the task axis.

Table 1. Average number of coaxial saccades in column-wise tasks, relative to the table-symbol and the table-grouping

	column-wise (s.d.)	neutral (s.d.)	row-wise (s.d.)
black-and-white	1.00 (0.60)	1.64 (0.95)	3.48 (2.36)
T-and-F	5.01 (3.13)	9.49 (4.18)	15.68 (5.84)

Table 2. Average number of coaxial saccades in row-wise tasks, relative to the table symbol and the table grouping

	column-wise (s.d.)	neutral (s.d.)	row-wise (s.d.)
black-and-white	4.49 (3.21)	2.77 (2.89)	1.19 (1.59)
T-and-F	13.71 (6.57)	9.31 (4.81)	4.23 (2.35)

5.3 Response Accuracy

Average percentages of correct answers are shown in Table 5, relative to the task axis, the table symbol, and the table grouping.

A repeated-measure 2×2×3 analysis of variance was conducted. We found strong effect of the table symbol ($F(1, 41) = 7.46, p < .01$), with performance with black-and-white tables being higher than performance with T-and-F tables. The main effect of the table grouping was also significant ($F(2, 82) = 3.72, p < .05$), indicating lower performance with column-wise tables compared to that with the other types of tables. The table grouping interacted with the task axis ($F(2, 82) = 4.41, p < .05$), where performance was better when the table grouping matched with the task axis than when they mismatched.

5.4 Response Time

The tendency of response time largely inherits that of response accuracy. Average response times are shown in Table 6, relative to the task axis, the table symbol, and the table grouping.

A repeated-measure 2×2×3 analysis of variance showed strong effect of the table symbol ($F(1, 41) = 318.1, p < .001$), where response time with black-and-white tables was shorter than response time with T-and-F tables. The main effect of the table

Table 3. Average number of backward saccades in column-wise tasks, relative to the table symbol and the table grouping

symbol	column-wise (s.d.)	neutral (s.d.)	row-wise (s.d.)
black-and-white	1.00 (0.37)	0.65 (0.39)	0.44 (0.22)
T-and-F	0.81 (0.50)	0.50 (0.42)	0.41 (0.39)

Table 4. Average number of backward saccades in row-wise tasks, relative to the table symbol and the table grouping)

symbol	column-wise (s.d.)	neutral (s.d.)	row-wise (s.d.)
black-and-white	0.51 (0.32)	0.84 (0.84)	0.91 (0.48)
T-and-F	0.53 (0.52)	0.86 (0.65)	0.71 (0.44)

Table 5. Average percentages of correct answers, relative to the task axis, the table symbol, and the table grouping

symbol	task axis	column-wise (s.d.)	neutral (s.d.)	row-wise (s.d.)
black-and-white	column-wise	0.94 (0.10)	0.95 (0.07)	0.93 (0.08)
	row-wise	0.91 (0.12)	0.96 (0.09)	0.96 (0.08)
T-and-F	column-wise	0.91 (0.11)	0.92 (0.11)	0.90 (0.11)
	row-wise	0.88 (0.15)	0.89 (0.14)	0.93 (0.11)

Table 6. Average response time (in milliseconds), relative to the task axis, the table symbol, and the table grouping

symbol	task axis	column-wise (s.d.)	neutral (s.d.)	row-wise (s.d.)
black-and-white	column-wise	2404 (415)	2671 (549)	2923 (645)
	row-wise	2938 (746)	2830 (563)	2554 (552)
T-and-F	column-wise	5135 (1315)	6563 (1559)	7225 (1884)
	row-wise	6499 (2092)	6195 (1481)	5158 (1331)

grouping was also strong ($F(2,82) = 20.09, p < .001$), indicating slower response with column-wise tables compared the other types of tables. The table grouping interacted with the task axis ($F(2,82) = 100.53, p < .001$), where response time was shorter when the table grouping matched with the task axis than when they mismatched. We also found strong interaction of the table symbol and the task axis ($F(1,41) = 13,07, p < .001$), of the table symbol and the table grouping ($F(2,82) = 9.26, p < .001$). The interaction of all the three factors was also significant ($F(2,82) = 49.51, p < .001$).

6 Evidences for Higher-Order Processing in the Comprehension of Spatial Representations

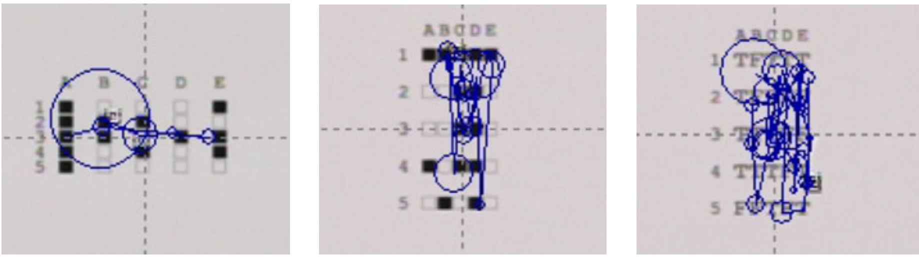
6.1 Attention to Large-Scale Objects

We predicted that fewer coaxial saccades would be found when the grouping of the given table matches with the axis of the task than when they mismatch. This prediction was clearly supported by our data. As Tables 1 and 2 show, coaxial saccades in column-wise tasks were less frequent with column-wise tables than with row-wise tables, whereas coaxial saccades in row-wise tasks were less frequent with row-wise tables than with column-wise tables.

Particularly, the average numbers of coaxial saccades were very small with black-and-white tables under the matching condition. For column-wise tasks, the relevant number was 1.00, and for row-wise tasks, the number was 1.19. Considering each table had five columns to be scanned, this means that, on average, approximately 0.2 scanning saccades were made on each column or row. Thus, our data for black-and-white tables confirm our prediction that the number of coaxial saccades in the matching condition would approach zero.

Figure 4 (1) shows the typical fixation pattern on column-wise, black-and-white tables during the column-wise task. The center of each circle indicates a fixation point, while its diameter indicates the fixation duration. Fairly long fixations were placed on individual columns, but eyes tended to leave a column after no or few vertical movements within the column. Rather, eyes moved from columns to columns consecutively, often placing only one fixation on each column.

By way of contrast, Figure 4 (2) shows the typical fixation pattern in the case where a column-wise task was performed using black-and-white, row-wise tables. We see more vertical saccades scanning columns of the table (3.48 on average). Figure 4 (3) shows



(1) Table: column-wise
Task: column
Symbol: black/white

(2) Table: row-wise
Task: column
Symbol: black/white

(3) Table: row-wise
Task: column
Symbol: T/F

Fig. 4. Typical fixation patterns observed in the experiment

the typical fixation pattern in the case where a column-wise task was performed using T-and-F, row-wise tables. Even more vertical saccades were observed in this case. (15.7 on average).

Thus, our data on coaxial saccades clearly support our hypothesis on the way columns or rows of symbols are bounded into coherent objects. For example, when column-wise tasks were performed with row-wise tables (Figures 4 (2) and (3)), initial attentions were oriented to consecutive sub-regions of a column, each consisting of one or a few symbols. Due to the task demand discussed in section 3, these initially attended sub-regions were then integrated into a column of five symbols, to which a task-relevant judgment was attached. Our data on coaxial saccades indicate that this type of subsequent integration took place generally under the mismatching conditions.

In contrast, when the table grouping matches with the task grouping, only one fixation was placed on an entire row or column to be scanned (Figures 4 (1)). This suggests that attention placed on one part of a row or column spread over the entire row or column, integrating five symbols into an object already at the time of initial attention. Subsequent integration was not necessary in such a case, and task-relevant judgment could be directly attached to the initially attended object. This could be enormous simplification of the relevant comprehension task, and explains shorter response time and higher response accuracy under the matching conditions compared to the mismatching conditions.

Ratwani et al [21] examined the operation of bounding large-scale objects in choropleth graphs on the basis of eye-tracking data. They observed that segmentation of small-size objects into a "visual cluster" was enabled by multiple fixations on the boundary of the cluster. Our finding is significant in suggesting that when the cluster is spatially coherent under the given task, the bounding can be accomplished through the spread of attention starting from one part of the cluster, as opposed to the multiple shifts of attention to its boundary.

6.2 Retainment of Large-Scale Objects

Our second hypothesis states that large-scale objects, once bounded, can be retained as coherent objects. This implies that backward saccades to large-scale objects can occur

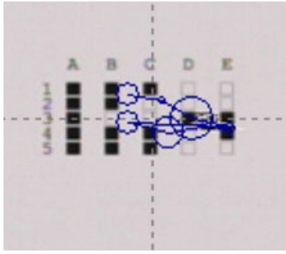


Fig. 5. Typical fixation patterns observed in the experiment

selectively. Furthermore, as the retainment may not be possible when the target does not make a good perceptual unit, the frequency of backward saccades should depend on the matching of the table grouping with the task axis.

Our data on backward saccades confirm both predictions. Backward conditions did occur in all conditions, the average 0.41 times per trial (twice every five trials) in the least frequent condition. The frequency of a backward saccade was the average 1.00 times per trial even in the most frequent condition, and the variance was not large (s.d. = 0.37). This indicates that background saccades in our data were *selective*, in the sense that they were directed to specific columns or rows, rather than directed to all the columns or rows indiscriminately. Figure 5 shows the fixation pattern of one of the trials where a single backward saccade occurred to a particular column in the middle of the trial.

Also, our data show that such backward saccades were more frequent when the table grouping matched with the task axis. The matching condition differed from the mismatching condition in that large-scale objects to returned to made good perceptual units. Thus, rows or columns with spatial coherence apparently facilitated returning saccades to them, and one good explanation for this facilitation is that those rows or columns were be retained in memory, as coherent objects for which perceptual reintegration out of individual symbols are not necessary. Under the mismatching condition, rows or columns lacked spatial coherence and were not retained as coherent objects in memory. Returning saccades to them were suppressed as a result.

In fact, the response accuracy was better in the matching condition than in the mismatching condition. Thus, it appears that by retaining large-scale objects in memory, people could effectively move their eyes back and forth, rechecking task-relevant information retrievable from those objects. Cognitive processes involving higher-level information were apparently facilitated by perceptual processes involving actual eye movements. This observation is in agreement with Ratwani et al [21], where "cognitive integration" of information obtained from visual clusters were evidenced by eye movements from clusters to clusters.

Retainment of large-scale objects in memory is a realistic possibility given the visual indexing mechanism investigated by Pylyshyn [19,20,1,29]. According to the visual indexing theory, we can assign "indices" to several objects or locations in the visual scene. With these indices, we can quickly return attention to the locations of indexed objects without searching for them. We hypothesize that such an index was attached to a row or column of symbols when it was first bounded. The index was then used for the quick return of attention when the maintenance need described above arose. Indeed, the use of eye movements for checking internally attached information was found also by Shimojima and Katagiri [28,27], and the present case is another instance of eye movements used in combination with visual indices. The present case is unique in that object groups, rather than individual objects or locations, were indexed and revisited for the task-relevant information.

However, retainment of higher-level objects is not the only possible explanation for the increased backward saccades and the better response accuracy in the matching condition. It might be that the spatial coherence of the relevant rows or columns made them perceptually more salient, so that they had been more easily re-identified (not traced back in memory) when backward saccades occurred. This could result in more frequent re-checking of the task-relevant information for those objects, making response accuracy better. More detailed analysis of the nature of backward saccades would be needed to clarify whether they constitute the re-identification of relevant rows or columns or the retracing of visual indices placed on them.

6.3 Further Questions

Although we found that rows or columns of tables can be bounded by a single shift of attention to them, our analysis is limited in that it examines only rows or columns of particular sizes. Thus, the question remains whether the same holds when stimulus tables are enlarged while the vertical and the horizontal spacing of the tables were preserved. Imagine, for example, making a table of pattern 2 (see Figure 3) twice as large both vertically and horizontally. Since this change does not affect the *ratio* of the vertical and the horizontal spacing of the table, the spatial coherence of columns of the table is not affected. The crucial question is whether Prediction (1) is true of such a case, namely, whether the number of coaxial saccades along a column of this enlarged table still approaches zero under the matching condition. As our data involve tables of pattern 2 only in a particular size, this remains an open question.

In particular, the length of a row or column in our stimulus tables was rather small under dense spacing. Recall, from section 3, that the threshold for maximum dispersion for a fixation was 1.67 deg in radius. This means that any eye movement within the distance of 1.67 deg from the center of gravity of previous eye samples is a minor movement within a fixation, rather than a saccade. As a row or column under dense spacing subtended only 4.0 deg, this in turn means that an entire row or column could fall inside the range of minor eye movements within a single fixation. Figure 6 shows this situation, where the large circle indicates the range of minor eye movements that could occur within a fixation placed on the position marked by the smaller circle.

Thus, strictly speaking, a large part of our results is only a confirmation of Prediction (1) as it is applied to the case where a large-scale object to be bounded is small enough to fall within the range of minor eye movements in a fixation. It says next to nothing about whether rows or columns substantially exceeding this range could be bounded by a single fixation. Several studies on object-based attention including [24] have reported the spread of attention over extended objects under eye-movement suppression, suggesting a positive answer to this question. An experiment that systematically controls the size of tables would be needed for a firmer answer.

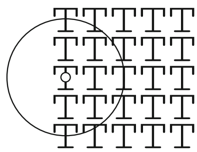


Fig. 6. Dispersion threshold for fixation identification, imposed over a stimulus table with dense spacing. The large circle indicates the range of eye minor eye movements that could occur within a fixation placed on the position marked by the smaller circle.

7 Conclusions

We argued that in order to fully understand how people utilize spatial representations for problem-solving, we should explicitly consider the issues of how people set the boundaries of graphical objects to be interpreted, and how the object boundaries are maintained during the task. Our experimental study with table-reading tasks supports the view that a large-scale object can be bounded by a single shift of attention, and that it can be retained as a coherent object for subsequent reference.

These findings in turn suggest a tight coupling of higher-level cognitive process and perceptual process in two important respects. First, the process of extracting higher-level information from a spatial representation depends on how easily the relevant object can be perceptually integrated, and made available to higher-level cognitive process: to extract higher-level information demanded by the given task, some spatial layouts (e.g., columns in column-wise tables) require only a single shift of attention to an object, while other layouts (e.g., rows in column-wise tables) require multiple shifts of attention. Second, the process of manipulating higher-level information thus obtained can be facilitated by perceptual-motor activities involving eye movements: to recheck higher-level information, some spatial layouts (e.g., column-wise grouping under a column-wise task) allow us to exploit retained large-scale objects by directing eyes back to them, while other layouts (e.g., row-wise grouping under a column-wise task) do not.

Overall, the study suggests a fundamentally integrative view, where the structures of higher cognitive processes can change *ad hoc*, in response to how graphical elements in the given spatial representation are managed in attentional process. Thus, subtle difference in spatial arrangement and visual properties of graphical elements can make structural difference to higher cognitive processes. In this regard, problem solving with spatial representations seems to be a more plastic process, deeply situated in the external world.

References

1. Ballard, D.H., Hayhoe, M.M., Pook, P.K., Rao, R.P.N.: Deictic codes for the embodiment of cognition. Behavioral and Brain Sciences 20(4), 723–767 (2001)
2. Barwise, J., Etchemendy, J.: Visual information and valid reasoning. In: Allwein, G., Barwise, J. (eds.) Logical Reasoning with Diagrams, pp. 3–25. Oxford University Press, Oxford (1990)

3. Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. The University of Wisconsin Press (1973)
4. Bertin, J.: Graphics and Graphic Information. Walter de Gruyter, Berlin (1981) (Originally published in France in 1977)
5. Blignaut, P.: Fixation identification: The optimum threshold for a dispersion algorithm. Attention, Perception, & Psychophysics 71(4), 881–895 (2009)
6. Driver, J., Baylis, G.C.: Attention and visual object segmentation. In: Driver, J., Baylis, G.C., Parasuraman, R. (eds.) The Attentive Brain, pp. 299–325. MIT Press, Cambridge (1998)
7. Duncan, J.: Selective attention and the organization of visual information. Journal of Experimental Psychology: General 113(4), 501–517 (1984)
8. Gilhooly, K.J., Wood, M., Kinnear, P.R., Green, C.: Skill in map reading and memory for maps. Quarterly Journal of Experimental Psychology 40A, 87–107 (1988)
9. Grawemeyer, B., Cox, R.: The effects of users' background diagram knowledge and task characteristics upon information display selection. In: Stapleton, G., Howse, J., Lee, J. (eds.) Diagrams 2008. LNCS (LNAI), vol. 5223, pp. 321–334. Springer, Heidelberg (2008)
10. Guthrie, J.T., Weber, S., Kimmerly, N.: Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations. Contemporary Educational Psychology 18, 186–221 (1993)
11. Kahneman, D., Treisman, A., Gibbs, B.J.: The reviewing of object files: Object-speci?c integration of information. Cognitive Psychology 24(2), 175–219 (1992)
12. Kosslyn, S.M.: Elements of Graph Design. W. H. Freeman and Company, New York (1994)
13. Kramer, A.F., Jacobson, A.: Perceptual organization and focused attention: the role of objects and proximity in visual processing. Perception & Psychophysics 50(3), 267–284 (1991)
14. Lamme, V.A.F., Rodriguez-Rodriguez, V., Spekreojse, H.: Separate processing dynamics for texture elements, bundaries and surfaces in primary visual cortex of the macaque monkey. Cerebral Cortex 9, 406–413 (1999)
15. Lowe, R.K.: Diagram prediction and higher order structures in mental representation. Research in Science Education 24, 208–216 (1994)
16. O'Craven, K.M., Downing, P.E., Kanwisher, N.K.: fMRI evidence for objects as the units of attentional selection. Nature 401, 584–587 (1999)
17. Olivier, P.: Diagrammatic reasoning: an artificial intelligence perspective. In: Blackwell, A.F. (ed.) Thinking with Diagrams, pp. 63–78. Kluwer Academic Publishers, Dordrecht (2001)
18. Pinker, S.: A theory of graph comprehension. In: Aritificial Intelligence and the Future of Testing, pp. 73–126. L. Erlbaum Associates, Mahwah (1990)
19. Pylyshyn, Z.: The role of location indexes in spatial perception: A sketch of the finst spatial-index model. Cognition 32, 65–97 (1989)
20. Pylyshyn, Z.: Seeing and Visualizing: It's Not What You Think. The MIT Press, Cambridge (2003)
21. Ratwani, R.M., Trafton, J.G., Boehm-Davis, D.A.: Thinking graphically: Connecting vision and cognition during graph comprehension. Journal of Experimental Psychology: Applied 14(1), 36–49 (2008)
22. Roelfsema, P.R.: Cortical algorithms for perceptual grouping. Annual Review of Neuroscience 29, 203–227 (2006)
23. Roelfsema, P.R., Lamme, V.A.F., Spekreijse, H.: The implementation of visual routines. Vision Research 40, 1385–1411 (2000)
24. Roelfsema, P.R., Lamme, V.A.F., Spekreojse, H.: Object-based attention in the primary visual cortex of the macaque monkey. Nature 395, 376–381 (1998)
25. Shimojima, A.: Operational constraints in diagrammatic reasoning. In: Barwise, J., Allwein, G. (eds.) Logical Reasoning with Diagrams, pp. 27–48. Oxford University Press, Oxford (1995)

26. Shimojima, A.: Derivative meaning in graphical representations. In: Proceedings of the 1999 IEEE Symposium on Visual Languages, pp. 212–219. IEEE Computer Society, Los Alamitos (1999)
27. Shimojima, A., Katagiri, Y.: An eye-tracking study of exploitations of spatial constraints in diagrammatic reasoning. In: Stapleton, G., Howse, J., Lee, J. (eds.) Diagrams 2008. LNCS (LNAI), vol. 5223, pp. 74–88. Springer, Heidelberg (2008)
28. Shimojima, A., Katagiri, Y.: Hypothetical drawing in embodied spatial reasoning. In: Proceedings of the 30th Annual Meeting of the Cognitive Science Society, pp. 2247–2252 (2008)
29. Spivey, M.J., Richardson, D.C., Fitneva, S.A.: Thinking outside the brain: Spatial indices to visual and linguistic information. In: Henderson, J.M., Ferreira, F. (eds.) The Interface of Language, Vision, and Action: Eye Movements and the Visual World, pp. 161–189. Psychology Press, San Diego (2004)
30. Stenning, K., Oberlander, J.: A cognitive theory of graphical and linguistic reasoning: Logic and implementation. Cognitive Science 19(1), 97–140 (1995)
31. Ullman, S.: Visual routines. Cognition 18, 97–159 (1984)
32. Wainer, H.: Understanding graphs and tables. Educaional Researcher 21, 14–23 (1992)

Enriching Spatial Knowledge through a Multiattribute Locational System

Stephen C. Hirtle and Samvith Srinivas

School of Information Sciences, 135 N. Bellefield Ave., University of Pittsburgh, Pittsburgh, PA 15260 USA
{hirtle,sas29}@pitt.edu

Abstract. Research has demonstrated that current navigation systems, while useful as a wayfinding tool, do not support the acquisition of spatial knowledge of the environment. In contrast to portable navigation devices, participants in this study learned about the location of lesser known campus libraries through a desktop system, which presented yoked images, maps and verbal directions to the participants. Information included not only the location of the building that housed the library, but also internal directions within the building to the library itself. The results showed that confidence to find the library by memory was notably improved by use of the system. Finally, the role of images was examined by selectively removing images from a random subset of the libraries, which resulted in marked decrease in confidence.

1 Introduction

Wayfinding is becoming a common task and is often accompanied with a navigational device such as the commonplace GPS navigation systems found in cars. These systems are quickly moving from the dashboard of the car to handheld devices. Unfortunately, there is also increasing evidence that the use of such devices can impair the acquisition of spatial knowledge about the world [1]. Furthermore, in well-travelled areas one often needs only simple cues (it is just past the Student Union) that complement your existing knowledge. These cues might take several forms such as text, map or street level images. This paper explores the notion that a stand-alone multi-attribute location finding system can be used effectively by adding to one's existing spatial knowledge. The system provides yoked maps, verbal directions and images, for both interior and exterior views, for discovering resources on a college campus.

The system builds on the work of Tversky [2], who uses the term *cognitive collage*, in contrast to *cognitive map*, to describe a person's spatial knowledge. The collage is a collection of partial bits of multimedia knowledge. Inherent within this collage is the ability to extract slices of information sources, such as visual cues, route information or linguistic labels. The collage necessarily operates at multiple levels, allowing one, for example, to discuss and plan a route, using a highway system or neighborhood streets with equal ease.

Inherent in the notion of a cognitive collage is the use of images for accessing our spatial knowledge. In fact, images in navigation are becoming more commonplace

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 279–288, 2010.

with the introduction of Google StreetView and other image-based navigation systems. Ichikawa and Nakatani [3] showed improved driver performance if the driver could watch a video of the trip before attempting to navigate on his own. Chen, et al [4] showed even greater performance if the video was of variable speed, slowing down near a decision point, then speeding up in sections between decision points. Meijer et al [5] showed that egocentric spatial knowledge of the layout and routes through a virtual supermarket was improved by the use of photorealism. In another recent paper, Oomes [6] found that presenting elderly pedestrians maps with photographs of landmarks near decision points assisted their acquisition of spatial knowledge, as did Kaminoyama [7] for those with dementia. Together, all of these recent studies suggest that visual images which match the "ground truth" can facilitate the acquisition of spatial knowledge.

The project also builds on current research arguing that a sense of place is necessary for effortless wayfinding and navigation [8]. The extent to which the presentation of new spatial knowledge can enable active learning is receiving much attention at present [9]. For example, Ishikawa et al [10] recently reported that using a guidebook for navigation supports long-term memory acquisition for visited sites, while using a mobile device does not. In this light, Montello [11] and others have argued for careful consideration in the adoption of new technology, particularly any fail-safe technology that does not facilitate the cognitive process of spatial learning, which would be needed when the technology fails to operate properly.

This project takes on a different focus than the previous work just cited, which focused on route learning and spatial knowledge acquisition. Specifically, we examine the situation of finding locations which are situated among known locations. Examples of this scenario would include when a new restaurant is described as being around the corner from a familiar pub, or finding a new store, which just opened in a familiar shopping district.

For this study, the problem is more specific. Here, we look at directing a student on campus to find a library that they have never been to before, but that is surrounded by familiar locations. The problem is further complicated since, unlike the restaurant just described, the library might very well be located off the main floor or away from the main entrance, so the navigation does not stop when one reaches the building. To assist students to solve this problem, a locator system was built, which includes maps, images and textual descriptions and is described in Section 2. Sections 3 and 4 describe a small usability study to measure the advantage of adding images over the other two kinds of information. Finally, the results are discussed with respect to the development of cognitively-driven wayfinding aides.

2 LibLoc

The Library Locator (LibLoc) system is a web-based browser that was designed to locate satellite libraries on the University of Pittsburgh campus [12]. There are 15 small libraries on the campus and while most students would be familiar with some library locations, many library locations are not well known. Furthermore, many satellite libraries are located in isolated locations deep inside an academic building. The system consists of four frames as shown in Figure 1 and 2, based on earlier work of Masui, Minakuchi, Borden and Kashiwagi [13].

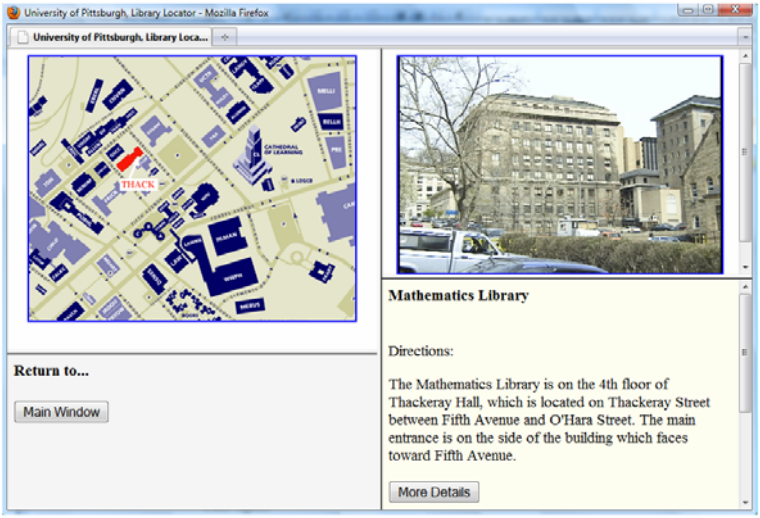


Fig. 1. Example screen from the Library Locator (LibLoc) showing an exterior map, image and text in a yoked system

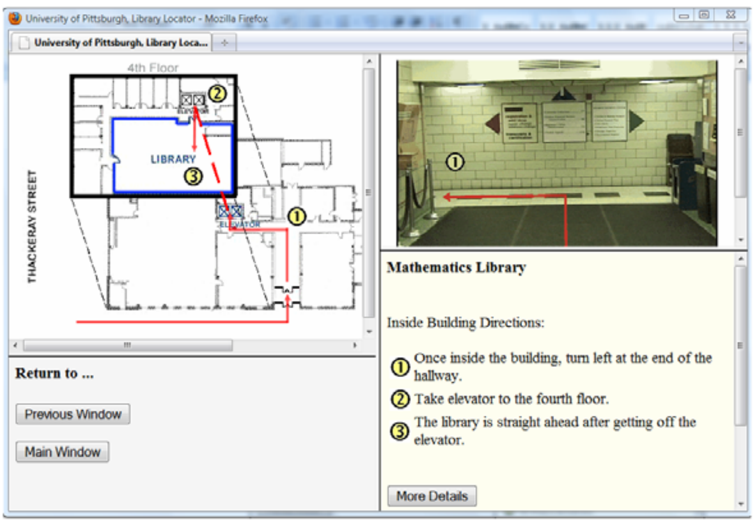


Fig. 2. Example screen from the Library Locator (LibLoc) showing an interior map, one of multiple images, and text in a yoked system

The upper-left quadrant is the spatial frame that contains a map of the campus, a 2D floor plan, or a 3D schematic of the building. The upper-right quadrant shows a key image or sequenced image, such as the front door of the target building or the inside of the library. The lower-right quadrant gives verbal instructions as to the location of the library and the lower-left window provides instructions for the use of the system. All information is presented in a structured, hierarchical manner, to allow users to explore within buildings, as well as around the campus, with equal ease.[1]

In this paper, we report the results of an empirical study in which participants used the system to learn about library locations. The design focused on the benefit of including images, so some locations were presented with images and some without. Participants were asked to rate their confidence. A small group of participants not only used the system, but where then asked to navigate to the actual locations. It is important to keep in mind that the system is designed to be used by existing students on campus in order to add to their spatial knowledge of the campus. While the system could be used by first-time visitors to the campus, it is more likely that naive users would benefit more from some kind of hand-held GPS device at the cost of not expanding their base spatial knowledge.

3 Method

3.1 Participants

A total of thirty students from the University of Pittsburgh were recruited through flyers posted around the University of Pittsburgh campus. Participants were paid $8 per hour for their participation in the experiment that lasted one to two hours. Their ages ranged from 20 to 47, with a mean of 26 years. The sample consisted of sixteen female and fourteen male participants.

3.2 Materials and Design

All participants were given a pre-test questionnaire that asked basic background questions followed by a 9-point rating of how familiar they were with the location of each of the 15 libraries on campus and a 5-point rating indicating how often they had visited each library in the past. The participants were then introduced to the LibLoc system, which was described in Section 2, and asked to systematically visit the eight lesser known libraries on the system. The location of the eight libraries varied in their complexity (low, medium, and high), as determined based on the structural complexity of the route (See Section 4.2 for details). For example, a low complexity direction needed one instructional unit to describe the location of the library, and a high complexity direction required six instructional units to describe the location of the library. For each participant, half of the libraries were shown with images and half were shown without images, counter-balanced across participants. Participants were instructed to study the web pages related to finding a particular library, in order to learn how to navigate to that location. Having used the system, 20 of the 30 participants (Study Only group) were asked to evaluate the system as well as rate their confidence in finding the locations. 10 of the 30 participants (Navigation group) were asked to travel across campus to the four randomly selected libraries to pick up a letter left at the circulation desk. Participants in the Navigation group then filled out a post-test questionnaire, which was similar to the Study Only post-test version.

---

[1] A similar approach can be seen in many of the online virtual campus tours (e.g., http://www.news.harvard.edu/tour/qtvr_tour/index.htm)

4 Results

4.1 Confidence

The twenty participants in the Study Only group were asked about their knowledge of libraries on campus, using a 9-point scale, where 1 indicated no knowledge of the location, before and after using LibLoc. Figure 3 shows the confidence of locating the eight lesser known libraries before and after using the system, with a small random scatter added to overlapping points to make them visible as a cloud pattern. For these eight relatively unknown libraries, ratings of the ability to find the library increase from an average of 3.79 to 7.53, t (159) = -19.13, $p < .01$ (alpha level .05; same in the analyses below) after using the system.

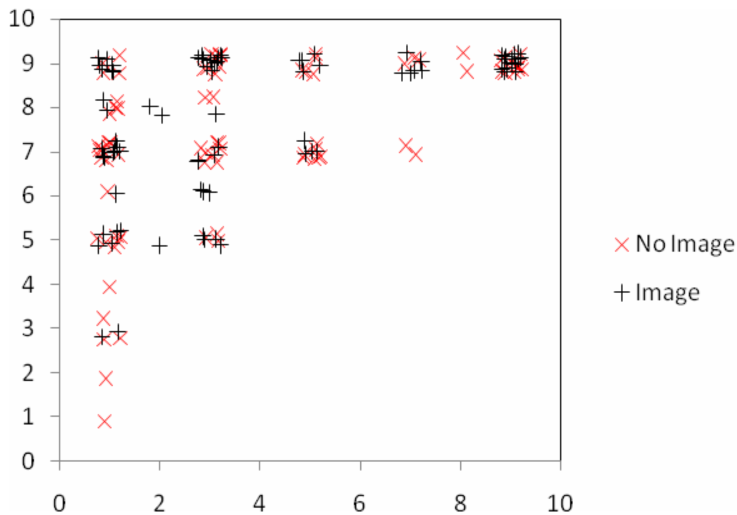


Fig. 3. Confidence before (horizontal axis) versus after (vertical axis) using the LibLoc system

While the eight libraries were chosen to be unknown, in fact there were 24 cases, out of the 160 pairs of judgments, in which participants were confident of the location at the start of the experiment and, not surprisingly, confident again at the end. Removing the 24 known libraries, shown as coordinates (9,9) in Figure 3, resulted in 136 judgments where some learning was possible. Of the 136 judgments, all but three cases, indicated an increase in confidence in locating the library after using the LibLoc system. Providing images resulted in a greater, but non-significant, increase in confidence (+4.64 for images versus +4.16 for no images, t = 1.36, p = .088).

4.2 Complexity

The eight key libraries, while all less known to the participants, also varied in the complexity of route within the building to find each library. This was measured as the sum of the number of turns plus the number of staircases plus the number elevators to reach the library. The overall complexity was then collapsed into one of three levels (Low, Medium or High), as summarized in Table 1. Table 1 also shows the

Table 1. Mean Complexity of Describing Library Location

Library	Complexity of Route*	Frames in Interior Slide Show	Number of Instructional Units	Overall Complexity
FFA	1	4	1	Low
GSPIA	1	5	1	Low
LL	1	5	1	Low
FLHS	2	6	2	Medium
PL	3	6	2	Medium
MaL	3	7	3	Medium
DML	6	9	6	High
MuL	7	10	6	High

* Complexity of Route = No. of Turns + Staircase + Elevator

number frames used in the interior slide show and the number of instructional units used in the verbal instructions, indicating that more information was needed to describe the more complex routes. The complexity measure could have been further refined by using the approach of Raubal and Egenhofer [14], who defined complexity in terms of the number of decision points and the strength of the wayfinding clues available at those points. However, given the collapse to three levels of complexity, both approaches would have yielded a similar classification.

Figure 4 compares confidence by complexity and whether or not images were included for that library. Two patterns emerge from the data. First, images results in an increase in confidence with an average improvement of ratings of 4.54 (with images) as compared with 4.21 (without images). Second, the highly complex locations resulted in a larger increase in confidence over the moderately and less complex locations. This may be a result of the participants having spent more time with complex locations.

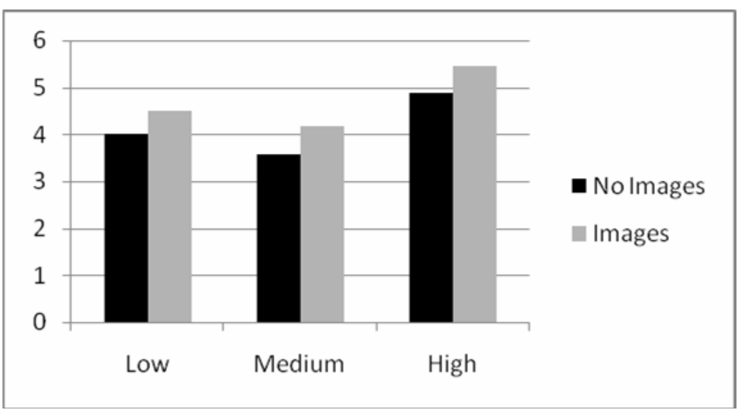


Fig. 4. Mean increase in reported confidence by level of complexity and inclusion of images

4.3 Navigation

As a confirmation of whether the confidence ratings were grounded in truth, ten additional participants (the Navigation group) used the system and then were set out to find all eight libraries on campus, in any order they wished. Participants were asked to locate each of the libraries, pick up a letter left at the circulation desk, and return to the location of the experiment after visiting all the libraries. This portion was self-paced and subjects were not timed nor tracked. All were able to easily find all the libraries without difficulty, despite not having a map or guidance system with them. Participants reported confidence levels that were similar to those reported in the study only group. For eight of the relatively unknown libraries, ratings of the ability to find the library increase from an average of 4.84 to 7.84, t (79) = -10.71, $p < .01$, after using the system.

4.4 Components

In terms of what components of LibLoc were most useful, participants were asked to identify what they thought was most important, second most important, and so on. Figure 5 shows the results, which indicate that while maps are a key feature to all participants, about half found verbal directions useful (first or second most important) and half found images useful (first or second most important). This no doubt reflects individual differences with regard to verbal and visual information, which is appropriate given that LibLoc is designed to facilitate those with different learning styles. This is consistent with the research of Mayer [15], who argues that superfluous images can distract from the learning process. In this study, images were not required for comprehension, but were made available to users who may recall seeing it before without knowing the name or departments housed there.

It is also worth noting that given the design of selective use of the images, a few participants voiced frustration with this part of the design, questioning why images did not occur with all the maps as they were quite helpful. This was due to the factorial experimental design and not a feature of the interface.

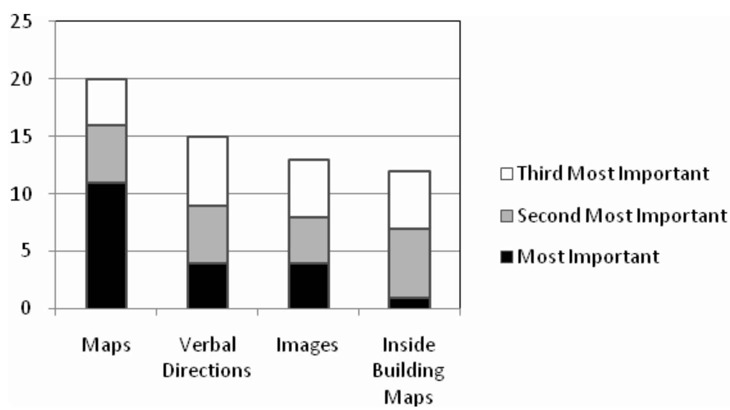


Fig. 5. Feature importance ratings for the Study Group

5 Discussion

The LibLoc system is designed to identify the knowledge base of the user and provide cues geared to that kind of user. It is a system that fills in the gaps of the user's knowledge, so that future wayfinding will not require a navigation system to find the location. The system supports individual differences by providing maps, images, and verbal directions to the user, and the results, in fact, indicated large differences in their preferred information source. In addition, inside building directions were provided, which even if not memorized, gave users a clue to the complexity of locating the building (take the elevator up to the 4^{th} floor, or take the stairs down to the basement). The inside routes also made clear why one entrance to the building was superior to another in cases where the library was near the entrance, which is often ignored by presenting only campus level maps.

The results overall showed that the system was successful in building confidence of the users as to the location of the library and that confidence was confirmed by a subset of users that made the trek across campus to find each of the targeted locations. Comments from users of the system included several usability issues, such as zoomable maps, but otherwise very positive comments such as

- "I really don't think anything about the system needs to be improved. I found it very helpful!"
- "I prefer the images, verbal directions and overall campus map working together to give me an intuitive whole view of these sites."
- "The interior pictures gave the greatest amount of assistance since most of the building I've been in at least once or twice. To actually see which staircase to take or which direction helped greatly"

The LibLoc system can be considered one example of how to increase the spatial knowledge of the user, while providing new spatial knowledge at the same time.

One's ability to navigate through space is multifaceted [16, 17]. Most automatic systems work extremely well and are becoming invaluable tools for travellers. However, the systems often lack the kinds of cues that would tie the directions to existing cognitive structures, such as local landmarks, neighborhoods, or visual cues, which is in contrast to human-generated directions [18, 19]. Yet, the ability to display images, maps, and text in dynamic displays allows one to provide complex directions with new clarity and precision that would augment the cognitive collage of the user [2, 20]. LibLoc provides one example of such a system.

This project was designed as a testbed for examining the role of maps, images and text, rather than a proposal for a functioning mobile system. That said, future work may lead to a mobile platform in which similar information is selectively displayed in concert with additional details of the library itself. This could be combined with recent work on geotagging images for navigation and identification [21] to provide the automated generation of a locational system. The goal of the of such a system would be not to simply arrive at a location, but to support the creation of a sense of place and spatial awareness [8].

There are other reasons to expand beyond considering navigation to be a simple task of getting from point A to point B. From a philosophical point of view, it is possible to consider the information content of a building [22], with history and context being encoded as part of the language of a building. From Borgmann's thesis, a building's content is not devoid of its material, construction, age, and so forth, but instead embraces its richness. To navigate around a major landmark, as if it were any other building in the space, denies the richness of the landscape. The graphic designers of the University of Pittsburgh understood the importance of landmarks by choosing to represent the Cathedral of Learning, a prominent 42-story building in the center of the campus, as the only building in three dimensions on the campus map. Portugali [23] has argued that such emergent properties of space should be considered in the design process, as the perception of space depends on the environment, which in turn directs further attention to the aspects of the environment that are perceived to be important. Good design can capture these qualities, but even in cases where design is lacking, information brokers, which include the image and context of the building, will lead to a great understanding of the space itself.

Acknowledgments

The authors thank Molly Sorrows, Stephen Hughes, Olena Shcherbinina and the other members of the Spatial Information Research Group (SIRG) at the University of Pittsburgh for their assistance in cultivating the ideas presented in this paper. The project also benefited from discussion of the experimental design with Christian Freksa at the Design Computing and Cognition ´04 Workshop on "Mental conceptions of emergent spaces in architectural design" at MIT.

References

1. Parush, A., Ahuvia, S., Erev, I.: Degradation in Spatial Knowledge Acquisition When Using Automatic Navigation Systems. In: Winter, S., Duckham, M., Kulik, L., Kuipers, B. (eds.) COSIT 2007. LNCS, vol. 4736, pp. 238–254. Springer, Heidelberg (2007)
2. Tversky, B.: Cognitive Maps, Cognitive Collages, and Spatial Mental Models. In: Campari, I., Frank, A.U. (eds.) COSIT 1993. LNCS, vol. 716. Springer, Heidelberg (1993)
3. Ichikawa, K., Nakatani, Y.: Driver Support System: Spatial Cognitive Ability and Its Application to Human Navigation. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4558, pp. 1013–1019. Springer, Heidelberg (2007)
4. Chen, B., Neubert, B., Ofek, E., Deussen, O., Cohen, M.F.: Integrated videos and maps for driving directions. In: 22nd Annual ACM Symposium on User interface Software and Technology, Victoria, BC, Canada, pp. 223–232 (2009)
5. Meijer, F., Geudeke, B.L., van den Broek, E.L.: Navigating through Virtual Environments: Visual Realism Improves Spatial Cognition. CyberPsychology & Behavior 12, 517–521 (2009)
6. Oomes, A.H., Bojic, M., Bazen, G.: Supporting Cognitive Collage Creation for Pedestrian Navigation. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics, HCII 2009. LNCS (LNAI), vol. 5639, pp. 111–119. Springer, Heidelberg (2009)
7. Kaminoyama, J., Matsuo, T., Hattori, F., Susami, K., Kuwahara, N., Abe, S.: Walk Navigation System Using Photographs for People with Dementia. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4558, pp. 1039–1049. Springer, Heidelberg (2007)

8. Klippel, A., Hirtle, S.C., Davies, C.: You-are-here maps: Creating Spatial Awareness through Map-Like Representations. Spatial Cognition and Computation 10, 82–93 (2010)
9. Ellard, C.: You Are Here: Why We Can Find Our Way to the Moon, But Get Lost in the Mall. Doubleday, New York (2009)
10. Ishikawa, T., Murasawa, K., Okabe, A.: Wayfinding and Art Viewing by Users of a Mobile System and a Guidebook. Journal of Location Based Services 3, 277–293 (2009)
11. Montello, D.R.: Cognitive Research in GIScience: Recent Achievements and Future Prospects. Geography Compass 3, 1824–1840 (2009)
12. Hirtle, S.C., Sorrows, M.E.: Designing a Multimodal Navigational Tool for Locating Buildings on a College Campus. Journal of Environmental Psychology 18, 265–276 (1998)
13. Masui, T., Minakuchi, M., Borden, G.R., Kashiwagi, K.: Multiple-View Approach for Smooth Information Retrieval. In: ACM Symposium on User Interface Software and Technology, Pittsburgh, PA (1995)
14. Raubal, M., Egenhofer, M.: Comparing the Complexity of Wayfinding Tasks in Built Environments. Environment and Planning B 25, 895–913 (1998)
15. Mayer, R.E.: Applying the Science of Learning Evidence-Based Principles for the Design of Multimedia Instruction. American Psychologist 83, 760–769 (2008)
16. Golledge, R.G.: Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes. Johns Hopkins Press, Baltimore (1999)
17. Taylor, H.A., Bruny, T.T., Taylor, S.T.: Spatial Mental Representation: Implications for Navigation System Design. Reviews of Human Factors and Ergonomics 4, 1–40 (2008)
18. Dale, R., Geldof, S., Prost, J.-P.: Generating More Natural Route Descriptions. In: 2002 Australasian Natural Language Processing Workshop, Canberra, Australia (2002)
19. Winter, S., Raubal, M., Nothegger, C.: Focalizing Measures of Salience for Route Directions. In: Meng, L., Zipf, A., Reichenbacher, T. (eds.) Map-Based Mobile Services - Theories, Methods and Design Implementations, pp. 127–142. Springer Geosciences, Berlin (2004)
20. Tufte, E.R.: Visual Explanation. Graphics Press, Cheshire (1997)
21. Bell, M., Reeves, S., Brown, B., Sherwood, S., MacMillan, D., Ferguson, J., Chalmers, M.: Eyespy: Supporting Navigation Through Play. ACM SIGCHI, 123–132 (2009)
22. Borgmann, A.: Holding on to Reality: The Nature of Information at the Turn of the Millennium. University of Chicago Press, Chicago (1999)
23. Portugali, J.: The Construction of Cognitive Maps. Kluwer, Dordrecht (1996)

Interactive Assistance for Tour Planning

Yohei Kurata

Department of Tourism Science, Tokyo Metropolitan University
1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan
ykurata@tmu.ac.jp

Abstract. It is often difficult for individual tourists to make a sightseeing tour plan because they do not have prior knowledge about the destination. Although several systems have been developed for assisting the user's tour planning, these systems lack interactivity, while demanding a lot of data input from the user. In this paper, we introduce a new computer-aided tour planning system, called *CT-Planner*, which realizes collaborative tour planning. The system provides several tour plans with different characters and asks the user to give feedback. The feedback is utilized by the system for inferring the user's preferences and then revising the tour plans. This cycle is repeated until the user is satisfied with the final plan. Thanks to this cycle the user does not have to register his profiles in advance. In addition, the system allows the user to specify his special requests, which leads to a more satisfying experience of computer-aided tour planning.

Keywords: spatial assistance system, tour planning, personalization, user's preference, human-computer interaction, selective travelling salesman problem.

1 Introduction

Thanks to the popularization of online reservation systems, it has become much easier for people to arrange their transportation and accommodation by themselves. However, it is still difficult for individual tourists to make a sightseeing tour plan, as they often do not have prior knowledge about the destination. Some guidebooks kindly show typical tour plans, but the problem is that attractive plans differ from person to person due to a large variety of tour preferences. Of course, this problem is not so serious when the destination is a small city or village where possible activities are limited. However, if people make a *short* visit to a large city with various *POIs* (*points of interest*), it becomes a very difficult problem for them to find an *efficient* tour plan (i.e., the plan that gives high satisfaction to the tourist in a short time) from a large possibility of tour plans.

To save individual tourists from such difficulty, several computer systems for supporting the user's tour planning have been developed. Some systems show a list of attractive POIs in the target area, taking the user's tour preference into account (e.g., [1-3]). However, even if people are informed about attractive POIs, it is still difficult for them to make a tour plan without the knowledge about the spatial arrangements of these POIs and the transportations between them. Thus, several systems were

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 289–302, 2010.

equipped with the ability to generate personalized tour plans [4-8]. Unfortunately, at the moment, these systems lack interactivity (e.g., we cannot request the system to insert certain POIs into the recommended tour plan), while demanding a lot of data input from the user (Section 2).

In this paper, we introduce a new computer-aided tour planning system, called *CT-Planner*, which overcomes the usability issues in the previous systems. *CT-Planner* stands for *Collaborative Tour Planner* (and also *City Tour Planner*). The key feature of this system is its *collaborative* planning process, which models the interaction between a tourist and a human tour advisor. First, the system provides several tour plans, which have different characters (nature-oriented, art-oriented, and so on). Shown these example plans, the user gives certain feedback to the system. From this feedback the system infers the user's preference and then revises the tour plans accordingly. By repeating this process, the recommended tour plan eventually agrees with the user's preference. This cyclic process nicely fits the assistance to ordinary tourists, because they usually do not have concrete requests from the very beginning, but rather the requirements are gradually formed during the process of planning [9]. This idea was briefly introduced in our previous work [10] among a number of ideas for future tourist information systems. Among these ideas, this paper focuses on the idea of the cyclic tour planning process. We concretize this idea in comparison with typical interactions between a tourist and a tour advisor, develops a prototype system that implements the idea, and demonstrates that this idea really works.

In the spatial sciences, many researchers studied problems in route planning (e.g., [11, 12]) and following navigation (e.g., [13, 14]). In contrast, the tour planning problem has attracted less interest, probably because it looks a domain-specific, application-oriented problem. Nevertheless, we believe the importance of this problem, because tourism is an activity that ordinary people have to handle heavy amount of spatial information to make decisions and they often need assistance from somebody. In addition, as an application of today's mobile and ubiquitous technologies, tourist assistance attracts more and more attention. In such a context, what we should pursue is appropriate assistance for the user's decision making, rather than simply providing more and more location-based information. We therefore consider that tour planning assistance will be one of the key topics for future research on spatial information systems.

The remainder of this paper is structured as follows: Section 2 gives a review of previous tour planning systems. Based on this review Section 3 discusses desirable designs of the systems. Section 4 introduces our implementation and its technical detail. Section 5 demonstrates how a user can make a tour plan in our system. Section 6 discusses the remaining issues toward more intelligent tour planning assistance. Finally, Section 7 concludes with a discussion of future problems.

2 Previous Tour Planning Systems

Most previous tour planning systems take similar three-step approaches for generating a tour plan. First, the system profiles the user's tour preference. Second, the system evaluates all POIs in the target area, considering the user's preference. And lastly, the system generates the 'best' tour plan, considering the calculated values of the POIs and given time/origin/goal constraints.

Profiling the user's preference is a problematic process. Early decision-support systems ask the user to set several preference parameters manually, for instance, by sliders (e.g., [4, 15]). This interface often frustrates the user, since he is forced to evaluate his own preference under given artificial criteria. To make the preference setting interface more user-friendly, Kurata *et al.* [5] introduced a questionnaire-based approach. In their system, the user is given fifteen questions, each showing two tour purposes and asking the user's preference on them (Fig. 1). Then, from the answers to these fifteen questions, the system calculates quantitative weights to ten tour purposes (Fig. 2) using AHP [16]. The problem of this approach is the time taken to answer fifteen questions [5]. To realize quicker user profiling, some systems ask the user to input his demographic data (age, gender, occupation, etc.) and calculates his preference parameters, assuming that tourists with demographically similar properties have similar tour preference [1, 8]. However, this approach also has a risk to annoy some users, giving an impression that the system offends their privacy and has a stereotyped view of their preferences. As an alternative approach, P-Tour [6] asks the user to evaluate the POIs by himself, from which the system generates the tour plan. This approach allows the user to specify his request directly and to skip the problematic profiling process, but it is hard work to estimate and input the attractiveness of all POIs in the target area.

Fig. 1. Questionnaire-based preference profiling in [5]

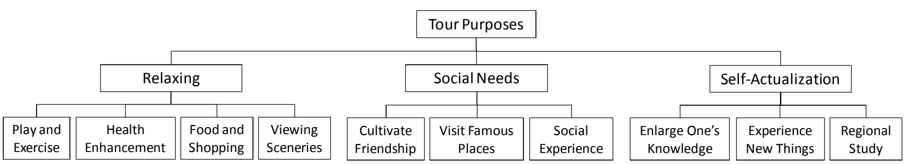


Fig. 2. Structure of ten tour purposes that forms the foundation of the user's preference model in [5]

Once the user's tour preference has been profiled, the system estimates the value of each POI for the user. The value may be calculated from the matching between the user's preference and the POI's chacter, or from the evaluation of this POI by other tourists with similar properties. In [5], each POI is scored by ten criteria, which correspond to the ten tour purposes in Fig. 2. Thus, by summing the scores of the ten criteria while applying the corresponding weights, the value of each POI for the user is calculated.

Finally, the system computes the 'best' tour plan. In the previous systems, the best tour plan is usually considered to be the plan that maximizes the total values of POIs to be visited during the tour. This problem is represented mathematically as follows (see also the illustration in Fig. 3):

Given a complete graph (V, E)*, the utility of each node* u_i*, the visitation time spent at each node* ${t^{visit}}_i$*, the travel time between two nodes* ${t^{travel}}_{ij}$*, origin node* $v_{ori} \in V$*, goal node* $v_{goal} \in V$*, and time constraint* T*, find a series of nodes to be visited* $v_{a_1}, \ldots, v_{a_k} \left(v_{a_i} \in V\right)$ *that maximize the sum of utilities* $\sum_{i=1}^{k} u_{a_i}$ *under the following three constraints:*

$$\sum_{i=1}^{k} {t^{visit}}_{a_i} + \sum_{i=0}^{k+1} {t^{travel}}_{a_i a_{i+1}} \leq T$$

$$v_{a_0} = v_{ori}$$

$$v_{a_{k+1}} = v_{goal}$$

Here each node represents a POI or a transportation node (e.g., a train station). The utility of a POI-node is given as the POI's value for the user, while the utility of a transportation node is set zero. The link and traveling time between each node pair may be computed beforehand in a GIS using the shortest path algorithm. On the other hand, visitation time spent at each POI-node has to be specified manually by experts. From a computational point of view, this problem is essentially the *Selective Travelling Salesman Problem* (*STSP*), which is known to be an NP-hard combinatory optimization problem [17]. Thus, the previous systems have adopted approximate methods, such as a genetic algorithm [7], to derive semi-optimal solutions in a practical time.

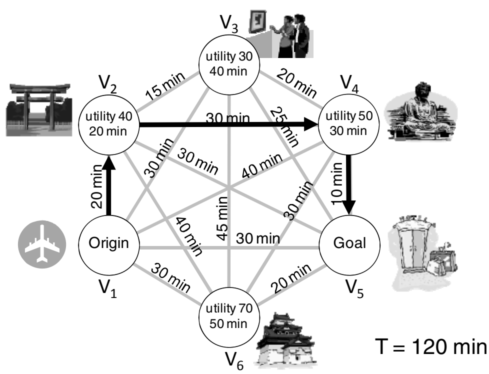


Fig. 3. Illustration of the choice of the 'best' tour plan

3 Interactive Tour Planning

Seifert [9] pointed out that the main disadvantage of personalized tour generation is that it excludes the user's participation in the process of planning. Indeed, according to the user test in [5], users complained about the inability to customize the recommended tour plans by adding or removing POIs which they want to visit/avoid. The tour planning systems in [1, 8] also have the same problem. Exceptionally, P-Tour [6] allows the user to express where he wants to visit/avoid by assigning high/low scores to the POIs. However, P-Tour forces the user to spend a lot of time to evaluate all POIs in the target area.

How can we avoid these problems? We can glean some hints from typical interactions between a tourist and a tour advisor. For instance, let us imagine the following conversation:

“I’d like to look around Yokohama in three hours or so.”

“Three hours. Okay, then, this is the most popular plan.”

“Well, hmm, looks like lots of museums in it. Honestly, I don’t have much interest in museums…”

“I see. Then, how about this plan? This visits more scenic sites along the port. Or, how about this plan visiting Sankei-en Park. You can enjoy lots of cherry trees in full bloom right now.”

“Hmm. For me, visiting portside scenic points sounds better. Well, but, I’d also like to visit the Port Museum. It’s very famous, isn’t it?”

“Yes, I definitely recommend it. So, how about this plan? First, you go …”

Usually a tour advisor does not know the interest of a tourist at the beginning of their conversation; the advisor learns the tourist’s interest from the conversation. The tourist’s response to recommendations indicates something about his preference. In many cases, the tourist is also not well aware of his interest at the beginning. He develops his interest while viewing the actual plans recommended by the advisor. Thus, experience advisor introduces several plans in a smart order, so that the tourist becomes aware of his own interest and able to judge what plan fits him.

Our system imitates this process; that is, we adopt the cycle where the system presents a set of plans to the user, infers the user’s preference from the user’s response, and then generates a new set of plans. This cycle for computer-assisted planning is called the *candidate/critique model* [18]. Although this cycle takes a bit of time to reach a final plan, the user probably attains a high level of satisfaction with the plan, as he can experience that his involvement directly results in the improvement of the plan. Probably the user would not attain such satisfaction, even if the smart system could derive the truly best solution in a single step. Another merit of this cycle is that the user does not have to specify his preference in advance. The user can discover his preference and requests through the comparison of actual plans.

Seifert [9] discussed a similar cyclic model of computer-aided tour planning. The difference from her approach is that our system *intentionally* generates several plans with different characters in order to seek the user’s tour preference, while her system simply generates the solutions—possibly many—that satisfy given constraints.

In the previous conversation, the tourist said that he did not have an interest in museums, while he requested to visit the Port Museum. Like this example, a tourist sometimes wants to visit a POI of the sort that he usually avoids. Conversely, he sometimes wants to skip a POI of the sort that he/she usually visits. For instance, a tourist, who usually likes modern art but not historical monuments, may request to visit the *Arc de Triomphe* because it is world-famous, or not to visit *Musée d'Orsay* because he has been there before. Computer-aided tour planning systems, therefore, should be able to support such case-by-case requests, just like a human tour advisor can do it.

4 CT-Planner

Based on the design concepts discussed in Section 3, we developed a new computer-aided tour planning system, called CT-Planner (Collaborative Tour Planner). As this name indicates, the system aims at the assistance of user-driven tour planning, rather than computer-based tour optimization. Currently the system is a stand-alone application for Windows PC, but we are also planning its server-client version that can be accessed from mobile devices.

Fig. 4 shows the main screen of CT-Planner. It shows two tour plans. The plan on the left is a recommended plan, which is computed based on the current model of the user's tour preference. The plan on the right is one of alternative plans. We have six alternative plans, namely *30min-shorter* plan, *30min-longer* plan, *more-education* plan, *more-art* plan, *more-nature* plan, and *more-experience* plan. The user can browse different alternatives by clicking tabs above the map and, if he wants, he can select the alternative plan that looks more attractive than the recommended plan shown on the left. Based on this selection, the time constraint or the model of the user's preference is updated and accordingly, all plans, including alternative plans, are revised under the new conditions. From the user's viewpoint, it looks that the alternative plan, if it is selected, moves to the left and becomes the next recommended plan. Note that we designed to display only two tour plans each time, following the key idea of AHP [16] that paired comparison is much easier than selecting one preferable plan from a number of alternatives.

If the user clicks the name tag of a node, its properties are shown in a new window (Fig. 5). In this window, the user can specify his request with regard to this node; that is, whether he wants to set this node as the origin/goal of the tour or, if it is a POI-node, whether he wants to visit/avoid this POI. The request yields the revision of all tour plans, including alternative plans. Once the user specifies a request, the system accommodates this request until he cancels it, no matter if he selects another alternative plan or changes the tour conditions afterwards.

Through repeated selection of preferable tour plans and/or specification of requests, the recommended plan will eventually become the plan that follows the user's preference and requests. Thus, when the user agrees with the recommended plan, the system prints out the plan and ends the process.

Our setting is similar to the setting of multi-criteria decision analysis (MCDA) [19], where the user is asked to select the best plan from multiple alternative. MCDA provides several techniques, such as AHP [16], to assign quantitative weights on given criteria. In our setting, education, art, nature, experience, and popularity correspond to the criteria in MCDA. Thus, we can alternatively apply MCDA techniques to decide the weights on five criteria. Actually, our previous system adopted AHP to calculate such weights [5]. We, however, did not apply AHP in the current CT-Planner, because under AHP the user has to compare the criteria themselves, which can be a burden for the user. Instead, we proposed a new technique to estimate the weights on the criteria indirectly through a cyclic process, in which the user can concentrate the comparison of actual plans.

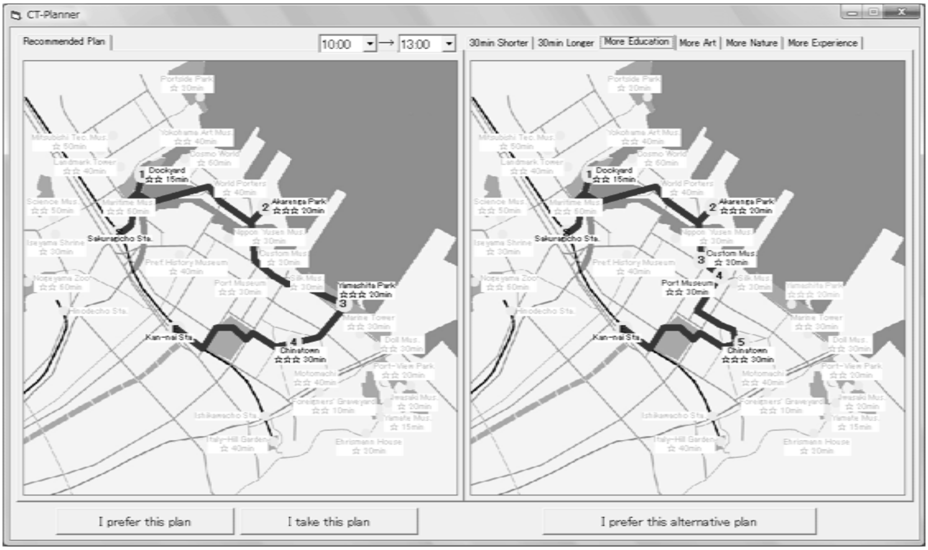


Fig. 4. The main screen of CT-Planner, which shows a recommended plan on the left and one of six alternative plans on the right

Fig. 5. A window showing the properties of a POI-node, in which the user can specify his request about this node

4.1 User's Preference Model

In CT-Planner, the user's tour preference is modeled by a five-dimensional unit vector (w_{edu}, w_{art}, w_{nat}, w_{exp}, w_{pop}), called the *preference vector*. Items w_{edu}, w_{art}, w_{nat}, and w_{exp} represent the user's weights on four types of tourist attractions—*education*, *art*, *nature*, and *experience*. This idea is similar to that of a *conceptual vector space* [20], in which instances of a concept are represented as points in a vector space. Raubal [20], for instance, represents building façades in a 7D space and considers the distance between instances to see the distinctiveness of each façade from others. Similarly, it is possible in our framework to evaluate the distinctiveness and similarity of tourists using our 5D conceptual vector space.

Initially, the weight is assigned only to popularity; i.e., (w_{edu}, w_{art}, w_{nat}, w_{exp}, w_{pop}) = (0, 0, 0, 0, 1). Naturally, the initial plan on the left becomes a 'standard' plan, which visits popular POIs as much as possible. If the user selects *more-education* plan, *more-art* plan, *more-nature* plan, or *more-experience* plan, the preference vector rotates toward the corresponding axis by θ degrees (Fig. 6). Such rotation is repeated if the user selects these plans again. In our current prototype, the parameter θ is initially assigned 45 degrees. If the user selects the recommended plan shown on the left, θ is reduced, because this selection means that all alternative plans are too biased for the user. Finally, if the user selects *30min-shorter* plan or *30min-longer* plan, the preference vector does not rotate.

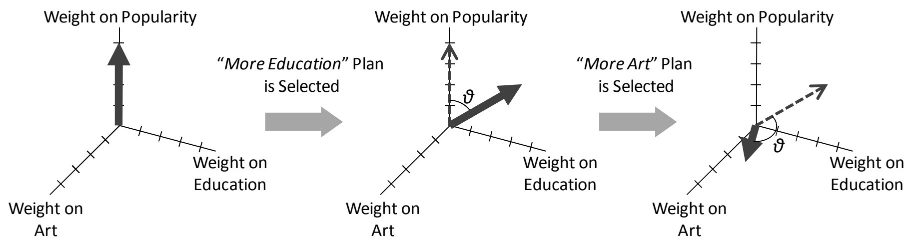


Fig. 6. Adaptation of user's preference model (here only three of five criteria are shown)

The reason why we adopted the user's weights on four types of tourist attractions, w_{edu}, w_{art}, w_{nat}, and w_{exp}, for representing the user's preference is that people often express their tour preference in terms of the type of tourist attractions they like. In this paper, we considered that *education*, *art*, *nature*, and *experience* are the four representative types of tourist attractions, which are highly independent from each other. There are other possible classifications—say, *nature*, *culture*, and *complex*, as used in the classification of UNESCO's world heritage sites, or *leisure* and *sightseeing* as used in many travel guidebooks. However, we should avoid detailed classification (say, more than five types), because under such classification we have to provide many tour plans for the user in each cycle, which would overwhelm the user (Section 4.3). Note that we further added one item w_{pop}, which represents the weight on 'popularity', because people typically prefer to visit popular POIs when they do not have any specific request.

4.2 Evaluation of POIs

Each POI has five-grade scores in five criteria—*satisfaction level of education*, *satisfaction level of art*, *satisfaction level of nature*, *satisfaction level of experience*, and *popularity*—which are determined by experts in advance. In other words, each POI is represented in another 5D conceptual vector space [20].

The value of the POI for each user is calculated from the total of these scores with the user's weight applied on the corresponding element. This value is calculated every time the user's preference model is updated.

On the screen (Fig. 4), the value of each POI is shown with one to three stars, instead of actual scores, such that the user easily figures out whether the POI is interesting to him or not.

4.3 Generation of Plans

For the generation of the recommended tour plan, we adopted the approximate algorithm by Kurata *et al.* [5]. This algorithm works in $O(nt^2)$, where n is the number of POIs in the target area and t is the length of tour time. With this algorithm, an ordinary PC generates a semi-optimal tour plan almost instantly. We believe that the optimality of the solution is not a serious problem in our application, since the plan is sufficient enough if it serves a nice draft, which is eventually customized by the user to satisfy his request in detail.

The six alternative plans are calculated by the same process but under slightly different conditions; that is, the *30min-longer* plan and 30*min-shorter* plans are the semi-optimal plans under the 30 minute longer/shorter time constraint, respectively, while the *more-education* plan, *more-art* plan, *more-nature* plan, and *more-experience* plans are the semi-optimal plans under the preference vector tentatively rotated toward the corresponding axis by θ degrees. As a consequence, if the user selects the alternative plan shown on the right, then this plan comes to the left side in the next step.

5 Demonstration

Let us observe how CT-Planner works in actual tour planning. Imagine a user who wants to plan a tour in *Yokohama* (a portside city near Tokyo) starting from *Sakuragi-cho Station* at 10:00 and arriving at *Kan-nai Station* before 13:00. Once the user inputs these basic conditions, the system generates two plans (Fig. 4). As stated before, the left plan is a popularity-based ‘standard’ plan, while the right plan is one of six alterative plans—for instance, *more-education* plan. The left plan is generated based on the preference vector (w_{edu}, w_{art}, w_{nat}, w_{exp}, w_{pop}) = (0, 0, 0, 0, 1), while the right one is based on (.707, 0, 0, 0, .707). In order to clarify the difference of these two plans, the POIs that only the alternative plan visits (in this case, *Custom Museum* and *Port Museum*) are given red name tags on the actual screen, while the POIs that the alternative plan misses (in this case, *Yamashita Park*) are given green name tags.

Imagine that the user selects the *more-education* plan, because he likes museums. Now the selected plan is transferred to the left side, while on the right side the system shows the second generation of *more-education* plan (Fig. 7a), which is generated based on the preference vector (1, 0, 0, 0, 0). He feels that this new plan contains too many museums and thus, he starts browsing other alternative plans. Then, he gets interested in *more-experience* plan (Fig. 7b) and selects it. Note that this plan emphasizes both education and experience, since the user has selected *more-education* plan in the previous step and accordingly, this plan is generated based on the preference vector (.5, 0, 0, .707, .5).

Now he notices that the current plan misses *Port Museum* where he wants to visit. Thus, he clicks the name tag of the *Port Museum* to open its property window (Fig. 5) and clicks the *Visit* button. Now the system shows a new plan that visits this museum

(Fig. 7c). However, due to the shortage of time, the new plan now skips *Akarenga Park*, which also looks attractive for him. Thus, he changes the arrival time from 13:00 to 13:30. As a consequence, the plan is revised, such that it contains both the *Akarenga Park* and the *Port Museum* (Fig. 7d). Since this plan look satisfactory for him, he clicks "*I take this tour*" button to print out it and finish planning.

In this way, CT-planner allows the user to make a tour plan in an interactive manner. Compared with the previous computer-aided tour planning systems, the user has much freedom to customize the recommended plans by himself. In addition, the user does not have to set preference parameters or input his demographic data for user profiling.

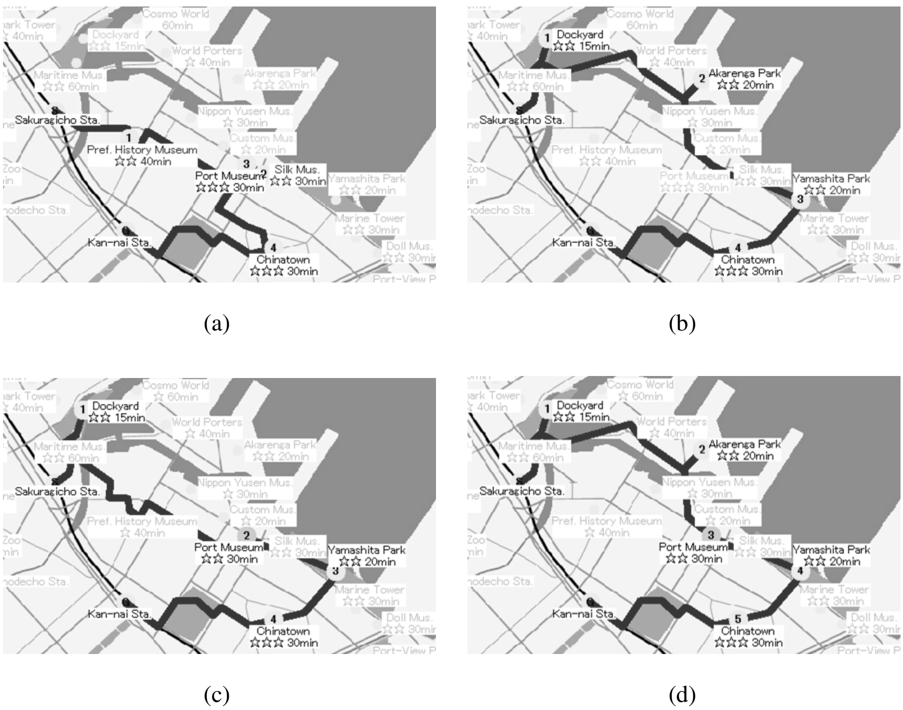


Fig. 7. Evolution of a tour plan in the scenario of Section 5

6 Toward More Intelligent Tour Planning Assistance

Although CT-Planner realizes easy and flexible tour planning, it has several remaining problems:

1. The attractiveness of routes is not considered.
2. Temporal/seasonal change of POI values is not supported.
3. Travel time between POIs is fixed, even though it may be affected by traffic jams, infrequent transportation service, or the user's choice of transportation modes.

4. Similarly, expected visitation time spent at each POI is fixed regardless of users.
5. Tour plans tend to become monotonous.

The followings are potential solutions to these problems, which revise the ideas discussed in our earlier paper [10].

An intuitive solution for Problem 1 is to assign values not only to nodes, but also to links. With this extension, the problem becomes the *Enhanced Profitable Tour Problem* (*EPTP*) [21]. The algorithm for deriving semi-optimal solutions of this problem is already proposed in [21].

Problem 2 is critical for actual tour planning. In the conversation in Section 3, for instance, the tour advisor told that the *Sankei-en Park* was nice to visit because its cherry trees were in full bloom. There are some POIs, like parks and botanical gardens, whose value changes by seasons. Similarly, there are some POIs, like scenic overlooks, whose value depends on the time of a day. An extreme example is museums, whose values are totally lost when they are closed at night. Thus, it is desirable that the system can support such temporal fluctuation of POI values. Matsuda *et al.* [22] already tackled Problem 2 (as well as Problem 3) from a computational point of view, introducing time-variant POI values and Fuzzified travel time between nodes.

As for Problem 4, a tourist probably spends more time at the POIs that fit his interest. Also, some people (perhaps older people) prefer slower tours. Thus, it is desirable that the system can estimate the time spent at each POI by each user, for instance, based on a certain statistical model.

Problem 5 arises from the fact that the current system evaluates POIs individually, but not in combinations. For instance, once the system learns that the user is interested in education, the system tends to recommend a tour plan that visits museums for the entire day—which may be boring even for museum lovers. Thus, it will be nice if the system can evaluate the *monotony* of each tour plan and utilize it for generating tour plans.

7 Conclusions and Future Work

For the people who travel unfamiliar places, tourist information is vital for their decision making. However, an overwhelming amount of information makes their decision rather difficult. Thus, how to customize the information for individual users is an important issue for tourist assistance systems, which should serve a wide variety of users. As an example of such systems, we introduce CT-Planner, with which people can make a tour plan easily as if they are assisted by a human advisor in a tourist information office. A nice feature of computer-aided tour planning systems, in comparison with human advisors, is that the user can consult the systems at any time, as much as he wants, using their own language even in foreign countries.

Our system is considered an example of *recommender systems*, which have been studied extensively in AI communities and is now used in many online shopping sites [23]. A major technique used in recommender systems is *collaborative filtering* (CF) [24], in which the system records each user's history of purchase or evaluation of items, finds similar users who have similar histories, and recommends an item that similar user has bought or given high evaluation. CF is particularly useful for the domains where it is difficult to provide explicit features of items and accordingly, it is difficult to model the user's preference (e.g., music and literature). Another merit is that each user does not have to report his/her preference to the system in an explicit manner. Meanwhile, in order to manage a practical CF-based service, we need a huge amount of past user data [23]. Unfortunately, our tour planning domain looks different from other domains where CF works well: First, the tourist's choice of a POI does not mean that he/she likes this POI. For instance, a tourist who has no interest in history may visit a historical site, in case this site is the only POI located between two POIs where he wants to visit. This implies a risk to consider that two tourists who have visited similar POIs have similar preferences. Second, even the same tourist may have different preferences at different times, depending on the destination, motivation, and companions, as found in the user survey of our previous system [5]. For instance, a tourist may prefer art-oriented tourism when she visits Paris by herself, while she may prefer nature-oriented tourism when she visits Switzerland with her husband. Thus, it is risky to rely on the tourist's preference model derived from his/her past behavior. For these reasons, we avoided a CF-based approach and adopted an approach using an explicit model of user's preference. Moreover, considering the instability of tourist's preferences, we decided not to store the profile of each user in the system.

We are planning a systematic user test in near future, which is necessary to substantiate our claim that CT-Planner gives more satisfying tour planning experience than the previous tour planning systems or guidebook-based planning. In this test, we also want to look at psychological aspects of tour planning processes in our user test, hopefully in collaboration with other researchers. For instance, we want to observe each user's interaction with the system—say, how many times he repeats the selection of alternatives and how much he customizes the recommended plan by himself—and seek the patterns of users' decision making. In order to conduct such survey, we are now equipping CT-Planner with the capability of recording the user's operations.

We are also planning to extend CT-planner, such that it works on mobile devices. There a key question is how to provide the service tailored to those devices. For instance, the mobile version of P-Tour [25] monitors the user's location and warns the user if he is off the route or behind schedule. Another idea is to modify the tour plan in case of sudden rain or delay. Such schedule monitoring and real-time rescheduling functions are potential strengths of mobile tour planning systems. In addition, mobile tour planning systems will become further useful if they are combined with other intelligent technologies for assisting tourists, such as route-specific route instructions [26] and location-based information querying by pointing [27].

Another potential strength of mobile tour planning systems is availability of trajectory data. Where a user visits and how much time he spends there should tell something about his preference [28]. Of course, a long stay in a certain POI does not necessarily indicate his interest in that kind of POI, because he may actually takes a rest at a café in the POI, for instance [2]. Thus, it is better to use microscopic trajectory data (hopefully both indoor and outdoor data) to analyze the user's action *in* a POI and seek his interest. The idea of such trajectory-based preference detection is also applicable to other spatial assistance systems. For instance, bike navigation systems may learn from the trajectory what kind of routes the user prefers. In shopping centers, data about where and how long the user spends time may be useful for adapting advertisements to the user. We therefore consider that trajectory-based preference detection is an important topic for the research on spatial assistance systems.

Acknowledgments

This work was supported by Tokyo Metropolitan University's research fund. We appreciate anonymous reviewers, participants of AGILE'09 Workshop on "Adaptation in Spatial Communication", and participants of GISA-Japan research conference 2009 for their precious comments and encouragement on our work.

References

1. Ricci, F., Arslan, B., Mirzadeh, N., Venturini, A.: ITR: A Case-Based Travel Advisory System. In: Craw, S., Preece, A.D. (eds.) ECCBR 2002. LNCS (LNAI), vol. 2416, pp. 613–627. Springer, Heidelberg (2002)
2. Schmidt-Belz, B., Nick, A., Poslad, S., Zipf, A.: Personalized and Location-Based Mobile Tourism Services. In: Workshop on Mobile Tourism Support Systems, pp. 18–20 (2002)
3. Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P.: Intrigue: Personalized Recommendation of Tourist Attractions for Desktop and Handset Devices. Applied Artificial Intelligence: Special Issue on Artificial Intelligence for Cultural Heritage and Digital Libraries 17, 687–714 (2003)
4. Kishimoto, H., Mizuno, S.: Development of Tour Planning Support System Using MDL and Genetic Algorithm. In: AI symposium 1997, SIG-KBS, pp. 71–76 (1997) (in Japanese)
5. Kurata, Y., Okunuki, K., Sadahiro, Y.: Development of a Preference-Based Tour Planning System. Papers and Proceedings of the Geographic Information Systems Association, vol. 9, pp. 199–202 (2000) (in Japanese)
6. Maruyama, A., Shibata, N., Murata, Y., Yasumoto, K., Ito, M.: A Personal Tourism Navigation System to Support Traveling Multiple Destinations with Time Restrictions. In: IEEE International Conference on Advanced Information Networking and Applications, vol. 2, pp. 18–22 (2004)
7. Goy, A., Magro, D.: Star: A Smart Tourist Agenda Recommender. In: Configuration Workshop on ECAI 2004, pp. 8/1–8/7 (2004)
8. Lee, J., Kang, E., Park, G.: Design and Implementation of a Tour Planning System for Telematics Users. In: Gervasi, O., Gavrilova, M.L. (eds.) ICCSA 2007, Part III. LNCS, vol. 4707, pp. 179–189. Springer, Heidelberg (2007)
9. Seifert, I.: Collaborative Assistance with Spatio-Temporal Planning Problems. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, D.R. (eds.) Spatial Cognition 2007. LNCS (LNAI), vol. 4387, pp. 90–106. Springer, Heidelberg (2007)
10. Kurata, Y.: Challenges in User-Adaptive Tour Planning Systems. In: Tomko, M., Richter, K.-F. (eds.) AGILE 2009 Workshop on Adaptation in Spatial Communication. SFB/TR 8 Spatial Cognition report series, pp. 19–26 (2009)
11. Mark, D.: Finding Simple Routes: Ease of Description as an Objective Function in Automated Route Selection. In: Second Symposium on Artificial Intelligence Applications, pp. 577–581. IEEE Press, Los Alamitos (1985)

12. Duckham, M., Kulik, L.: Simplest Paths: Automated Route Selection for Navigation. In: Kuhn, W., Worboys, M.F., Timpf, S. (eds.) COSIT 2003. LNCS, vol. 2825, pp. 169–185. Springer, Heidelberg (2003)
13. Denis, M.: The Description of Routes: A Cognitive Approach to the Production of Spatial Discourse. Current Psychology of Cognition 16, 409–458 (1997)
14. Klippel, A., Tappe, H., Kulik, L., Lee, P.: Wayfinding Choremes: A Language for Modeling Conceptual Route Knowledge. Journal of Visual Languages and Computing 16, 311–329 (2005)
15. Hochmair, H., Rinner, C.: Investigating the Need for Eliminatory Constraints in the User Interface of Bicycle Route Planners. In: Cohn, A.G., Mark, D.M. (eds.) COSIT 2005. LNCS, vol. 3693, pp. 49–66. Springer, Heidelberg (2005)
16. Saaty, T.: The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. McGraw-Hill, New York (1980)
17. Laporte, G., Martello, S.: The Selective Travelling Salesman Problem. Discrete Applied Mathematics 26, 193–207 (1990)
18. Linden, G., Hanks, S., Lesh, N.: Interactive Assessment of User Preference Models: The Automated Travel Assistant. In: Jameson, A., Paris, C., Tasso, C. (eds.) Sixth International Conference on User Modeling, pp. 67–78. Springer, Heidelberg (1997)
19. Malczewski, J.: GIS and Multicriteria Decision Analysis. Wiley, Chichester (1999)
20. Raubal, M.: Formalizing Conceptual Spaces. In: Varzi, A., Vieu, L. (eds.) FOIS 2004, vol. 114, pp. 153–164. IOS Press, Amsterdam (2004)
21. Joset, M., Stille, W.: A User-Aware Tour Proposal Framework Using a Hybrid Optimization Approach. In: Voisard, A., Chen, S. (eds.) 10th ACM International Symposium on Advances in Geographic Information Systems, pp. 81–87. ACM Press, New York (2002)
22. Matsuda, Y., Nakamura, M., Kang, D., Miyagi, H.: An Optimal Routing Problem for Sightseeing with Fuzzy Time-Varying Weights. IEEE International Conference on Systems, Man and Cybernetics 4, 3665–3669 (2004)
23. Bachrach, Y., Porat, E., Rosenschein, J.: Sketching Techinique for Collaborative Filtering. In: International Joint Conference on Artificial Intelligence, pp. 2016–2021. AAAI Press, Menlo Park (2009)
24. Rensnick, P., Iacovou, N., Suchak, M., Bergstorm, P., Riedl, J.: Grouplens: An Open Architecture for Collaborative Filtering of Netnews. In: ACM 1994 Conference on Computer Supported Cooperative Work, pp. 175–186. ACM Press, New York (1994)
25. Shiraishi, T., Nagata, M., Shibata, N., Murata, Y., Yasumoto, K., Ito, M.: A Personal Navigation System with Functions to Compose Tour Schedules Based on Multiple Conflicting Criteria. IPSJ Digital Courier 1, 528–536 (2005)
26. Richter, K.-F., Klippel, A., Freksa, C.: Shortest, Fastest, - but What Next? A Different Approach to Route Directions. In: Raubal, M., Sliwinski, A., Kuhn, W. (eds.) Münsteraner GI-Tagen 2004, pp. 205–217. IFGI prints, Münster (2004)
27. Frank, C.: An Egocentric Spatial Data Model for Intelligent Mobile Geographic Information Systems. Department of Spatial Information Science and Engineering, Master's Thesis. University of Maine, Orono, Maine, USA (2003)
28. Kanda, T., Shiomi, M., Perrin, L., Nomura, T., Ishiguro, H., Hagita, N.: Analysis of People Trajectories with Ubiquitous Sensors in a Science Museum. In: IEEE International Conference on Robotics and Automation, pp. 4846–4853 (2007)

Verbally Annotated Tactile Maps – Challenges and Approaches

Christian Graf

cgraf@acm.org

Abstract. Survey knowledge of spatial environments can be successfully conveyed by visual maps. For visually impaired people, tactile maps have been proposed as a substitute. The latter are hard to read and to understand. This paper proposes how the cognitive disadvantages can be compensated for by Verbally Annotated Tactile (VAT) maps. VAT maps combine two representational components: a verbal annotation system as a propositional component and a tactile map as a spatial component. It is argued that users will benefit from the cross-modal interaction of both. In a pilot study it is shown that using tactile You-Are-Here maps that only implement the spatial component is not optimal. I argue that some of the problems observed can be compensated for by incorporating verbal annotations. Research questions on cross-modal interaction in VAT maps are formulated that address the challenges that have to be overcome in order to benefit from propositional and spatial representations induced by VAT maps.

Keywords: verbal annotation, tactile map, representation, navigation, representational modality, multimodality.

1 Introduction

From the perspective of cognitive science, human knowledge and reasoning processes involve representations that serve as counterparts to real-world entities. There is strong evidence for an analogue nature of representation of spatial environments, i.e., they have intrinsic spatial properties [1]. Investigations into mental spatial representations suggest mental maps [2] and spatial mental models [3], have a spatial nature. These internal representations can be induced via linguistic or perceptual inputs from external representations.

External, diagrammatic representations, such as maps, can serve as a means for capturing and processing knowledge [4] and for solving spatial reasoning tasks [5]. Maps appear to be a promising aid for visually impaired people as well [6], although there are divergent results about spatial processes and representation of space in blindness [7, 8]. To accommodate the special abilities of users who cannot access visual information the substitution of visual perception by tactile perception allows the use of *tactile maps*. Such maps have become an option for supplying geographical knowledge to visually impaired people [9]. One class of maps, You-Are-Here (YAH) maps [10], can successfully support the task of navigating in complex in-door (for

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 303–318, 2010.

example, malls, hospitals, etc.) and out-door environments (such as, parks, zoos, university campuses, etc.). YAH maps are multi-purpose maps specifically made to provide an overview of the proximate surrounding of the geographical position that the map-user is situated in when reading the map. They have proven to be successful in facilitating wayfinding for sighted people [11]. By providing tactile YAH maps, visually impaired persons can be supported in the acquisition of *survey knowledge* [12]. Knowing about the structure of the surrounding environment could enable independent wayfinding. GPS-based systems, in contrast, are often made to provide the path to a specific locomotion or solution to a wayfinding problem.

There is strong evidence that such systems have negative effects on the acquisition of spatial knowledge [13]. By consulting a tactile YAH maps before a trip, a model of the environment could be conveyed, exploration behavior could be guided, and the visually- impaired persons' self-dependency could be supported, particularly in environments visited for the first time.

A first objective of the research reported in this paper is to test tactile YAH maps for their effectiveness in conveying the spatial layout of an environment. I highlight the properties of YAH maps and discuss how those properties may be transferred into the tactile domain with the objective of *cognitive adequacy* [14]. A cognitively adequate map is an external map that results in an internal mental representation that enables the map-reader to solve spatial reasoning tasks successfully. Individual parameters in the realization of tactile YAH maps will be investigated. This paper offers some ideas for how these parameters might convey spatial knowledge. The study highlights some limits on the usage of tactile YAH maps and supports the usage of *Verbally Annotated Tactile maps* (VAT maps) as extensions of tactile maps.

The concept of a *Verbally Annotated Tactile map* is introduced. This is a map that incorporates verbal content (of the map or about the map) during the time of proposed exploration. The concept is discussed in terms of representational theory; it is identified as a *multimodal representation* integrating *spatial* and *propositional representations*. VAT maps might provide solutions to the some current problems with tactile maps. Challenges with VAT maps are identified and a research agenda is described that points to some potential developments such as *Verbally Annotated Virtual Tactile Maps*.

In the next section I start with some background on tactile YAH maps, then I report an experiment about how well participants performed in a wayfinding task with a tactile map. Section 3 details the concept of VAT maps, presents some theoretical considerations about a verbal annotation system for tactile maps, and ends with elaborating on future research. Conclusions are offered in Section 4.

2 Tactile You-Are-Here Maps

YAH maps have an intermediary status [15]: They are not specialized like a route description, which provides instruction for one single route from *one* location to *one* destination. Furthermore, they are not general like a multi-purpose city-map, which can be used to navigate from *many* locations to *many* destinations. Instead a YAH map enables navigation from *one* location, where the map and the navigator are co-located (i.e., the You-Are-Here position), to *many* destinations. As the YAH map has to serve multiple purposes it usually displays details at the same level of granularity over the map (in contrast to, e.g., focus maps [16]). YAH maps are often found at junctions or other main decision points. They enable people to localize themselves in a depictive representation of the environment. After self-localization map-readers might engage in further exploration of the map for an overview of the area or they might start to search for specific routes. YAH maps have one aspect that is usually prominent so that it can be found quickly, the YAH symbol.

Transforming a visual YAH map into a tactile equivalent offers some challenges. In some circumstances the tactile sense can substitute for other sensory modalities like vision [17]. For a review of *sensory substitution* from visual to tactile media see [18]. But tactile maps are not equivalent to visual ones [19, 20]. First, as consequence of the biological properties of the receptors in the skin there is a limitation in resolution of what humans can sense by touch. A consideration of the physiology and the psychophysics of touch reveals a multitude of receptors that make different types of sensations possible, for example the sensations of temperature, deformation, and lateral movements [21]. By touch it is possible to discriminate objects using their contours, surface roughness, and material. To clearly feel one entity and separate two entities from each other, the entities and the distance between them have to be greater than in vision. As a consequence, the distribution of objects in effective tactile presentations must be sparser than in visual ones (see [22] for exact figures). Thus, a reduction in content is necessary to represent a scene with a tactile map. Alternatively, the size of the tactile map would have to be increased to an unmanageable size. Second, abstracting graphical entities in reduced-detail maps may lead to significant errors, for example misinterpretations. Michel showed that additional transformations like distortions and displacements could be useful to direct attention and to make a tactile map and its entities meaningful to users [23]. Third, using a tactile map can limit some cognitive processes. Instead of having an immediate visual impression of a map, the user has to serially explore the surface of a tactile map with his fingertips and integrate the properties and impressions of single points of contact into one compound impression of the map. Systematic training of appropriate tactile scanning strategies was found to be necessary to help blind children improve their ability to orient themselves with a tactile map [24].

The investigation into the feasibility of communicating spatial knowledge with tactile YAH maps to form an internal representation that can be used to solve spatial cognition tasks, serves as a baseline for further research into VAT maps. The investigation of a unimodal type of map (i.e., the tactile maps) should be a basis for and should inform further research about multimodal types of maps (i.e., the VAT maps). Here we report the results of a pilot study that addresses some of these basic issues.

2.1 Research Questions

In a simple form, the intuitive question is, how could tactile YAH maps be implemented with different tactile entities for the same conceptual function so that it is helpful? Specifically, is the map as usable as possible and at the same time useful? In this experiment it is assumed that YAH map-users want to locate themselves first before getting to know the environment. Therefore, they look for the YAH point – denoted by the YAH symbol –before exploring the map. Two concrete questions are considered:

1. Which (tactile) guide type for the You-Are-Here point in a tactile map is the most effective, i.e. has the shortest search time?
2. Which guide type is the least hindering when exploring the map to build up some survey knowledge in terms of objective time needed for exploration, objective quality of acquired survey knowledge, and subjective judgments of the map?

The comparison of the different measures should show if there is a mismatch in objective and subjective assessments, that is, whether people might subjectively regard one guide type better than another even if it is objectively not.

2.2 Experimental Design

The experiment had a within-subject design with one factor and three conditions. The independent variable was which guide type to the YAH symbol was used. The dependent variables that are reported here were the search time to find the YAH symbol, the exploration time, the time to draw a sketch, and the quality of a sketch.

1. Guiding Line Condition (GL): A unique line guided the user from a prominent starting point (marked with an arrow) on the frame of the map to the position of the YAH symbol in the map;
2. Grid Condition (GR): A grid of unique lines partitioned the map into regions that define the coordinates; the position of the YAH symbol was given in these coordinates;
3. Frame Marks Condition (FM): Four arrow-like marks, two on the vertical and horizontal borders of the map, defined the horizontal and vertical position of the YAH symbol in the map.

2.3 Materials

The tactile maps were in A4 format and produced with a ViewPlus® Emprint SpotDot™ Braille Embosser based on the TIGER® technology that embosses tactile pixels (following the convention in literature I call this entity a “taxel”) into a paper surface of size A4 (29.7 x 21 cm) with a resolution of 20 taxels per inch [25]. Assembling single taxels beside each other composes objects like (raised) lines, figures and regions. All tactile objects on the maps had the same surface structure and the same prominence in terms of height above the base material. The tactile maps were of uniform scale and each presented a grid of perpendicular tracks (4 horizontal ones and 4 vertical ones with a total of 12 intersections) with three landmarks (the YAH symbol plus representations of two buildings). The maps were of a virtual environment to control for unwanted effects of familiarity, so that participants had to learn all spatial relations from the map. The structure of tracks in the maps was the same but the geometry was different (distorted and rotated) between the maps. Landmarks were placed near tracks in such way that no two landmarks occupied the same region. The shortest paths between the landmarks were of uniform complexity (measured in number of turns/legs) in each map: from the YAH point to landmark 1 with a single turn, to landmark 2 with two turns, and between landmark 1 and 2 with three turns. For two examples see Figure 1.

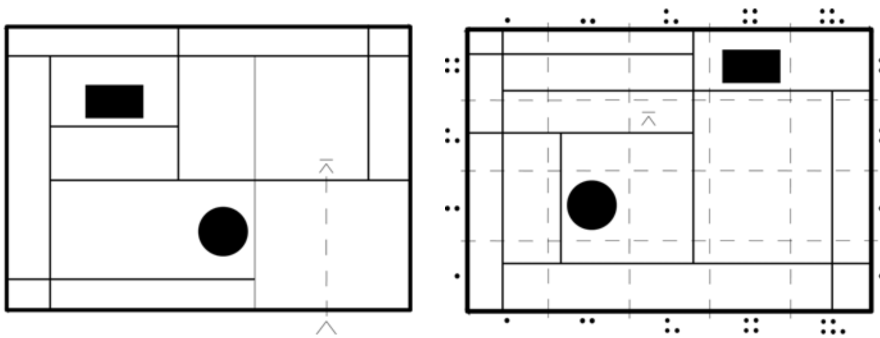


Fig. 1. Visualization of the tactile maps used as stimulus in the condition Guiding Line (left) and in the condition Grid (right)

2.4 Participants

The participants in all experiments reported here were 12 sighted individuals (all students, 4 male & 8 female, *M* = 28 years, *SD* = 5 years) with no impairment to the visual, tactile or motor system and no experience in solving task solely with their tactile sense. Sighted participants were chosen because they are familiar with maps and understand map concepts and conventions - many visually impaired persons have no experience with maps. The participants showed a reasonably high self-confidence in their abilities to read visual maps (*M* = 0.72, *SD* = 0.17, with normalized values) and to successfully solve tasks with visual maps (*M* = 0.79, *SD* = 0.07).

2.5 Procedure

The participants sat at a desk in front of a curtain to exclude direct sight of the tactile material behind the curtain. They put their hands and arms beneath the curtain to reach the tactile material. In the beginning of each session, participants were individually trained on sensing and interpreting tactile sensory input. Each tactile symbol that was used in the experiment was introduced in the training and participants had to pass a sensory recognition test before the experiment.

During the experiment, all participants handled the maps in all conditions. The order of the conditions was systematically varied between participants. In each condition before the tactile map was presented, participants explored a legend tactually and could ask about the meaning of symbols and lines. If requested a pre-defined explanation was read out. Instructions for how to use the map and how to find the YAH point were given verbally similar to what would normally occur if the legend and caption of a tactile map were given in Braille.

The first task in each condition was to find the YAH symbol in the map as quickly as possible. The time from touching the map until participants indicated they were sure they had found the YAH symbol was measured.

In a second task, the participants were asked to explore the map in a self-paced manner so that they would be able to explain routes and the structure of the environment without consulting the map. The participant's goal was to gain an understanding of the entire environment, without any specific route given beforehand. To check for a basic understanding of the environment, participants were asked questions after the exploration of a map: what buildings are in the environment, and how are those buildings distributed in the environment. If a participant could not answer both questions, they were allowed to further explore the map. The total time for all explorations was recorded.

In a third task, participants were asked to produce a complete copy of the on a graphic tablet to test their survey knowledge of the scene. No time limit was given, but time to complete the drawing was recorded.

A fourth task describing two routes between landmarks was given but is not reported here. The complete experiment lasted about two hours (M = 2:02:55h, SD = 15:06min).

2.6 Results

The efficiency of the different guide types was assessed using the time to locate the YAH symbol. All participants except from one in the Frame Marks condition completed the task. In the Frame Marks condition map users located the YAH symbol the fastest, and were slowest in the Grid condition (see Table 1). After two outliers were excluded[1], an ANOVA test revealed that the differences among condition reached the .0001 significance level, $F(2, 31) = 15.13$, $p = .000026$.

Concerning which guide type was least hindering in the exploration phase, a comparison of the exploration times (see Table 1) showed that in the Guiding Line conditions participants explored the maps fastest. An ANOVA test confirms a significant difference between all conditions at the .01 significance level, $F(2, 33) = 7.23$, $p = 0.0025$.

There was no significant effect of condition on time needed to sketch a map was found, $F(2, 32) = 0.097$, $p = 0.91$. A descriptive overview of search times, exploration times and drawing times is given in Table 1 below.

Table 1. The three main quantitative variables recorded during the experiment (in minutes)

	Time to find the YAH symbol	Time for exploring the map	Time for drawing the sketch	
GL	0:22	4:11	1:27	
FM	0:16	5:59	1:23	
GR	1:46	7:56	1:37	
	0:48	*6:02*	*1:27*	*Mean*
	0:50	*1:53*	*0:07*	*SD*

[1] The following convention is used: a data point that exceeds the range [(lower quartile - 1.5 * interquartile range) to (upper quartile + 1.5 * interquartile range)] is considered an outlier. See “What are outliers in the data?” In: *NIST/SEMATECH e-Handbook of Statistical Methods*, Chapter 7.1.6. http://www.itl.nist.gov/ div898/handbook/, retrieved 2010-04-12.

The sketches that the participants produced after the exploration of the tactile map (in total 12x3=36 sketches) were rated by two independent experts on, could a person that does not know the area navigate it successfully? Grade 1 was awarded for a perfect map that showed the landmarks and streets in correct relation to each other (although there might be minor flaws in geometric details), grade 6 was given for a map if there was no resemblance to the structure of the tactile map. In total 6 sketches were rated very good or good, 26 acceptable or moderate and only 4 not interpretable. Together 32 of 36 sketches (>88%) were at least of moderate quality and indicated that the map-readers probably did build a mental representation of the environment depicted in the tactile maps. The tactile maps effectively conveyed spatial knowledge to the participants in almost all cases. See Figure 2 for two examples of sketch maps produced by two different participants.

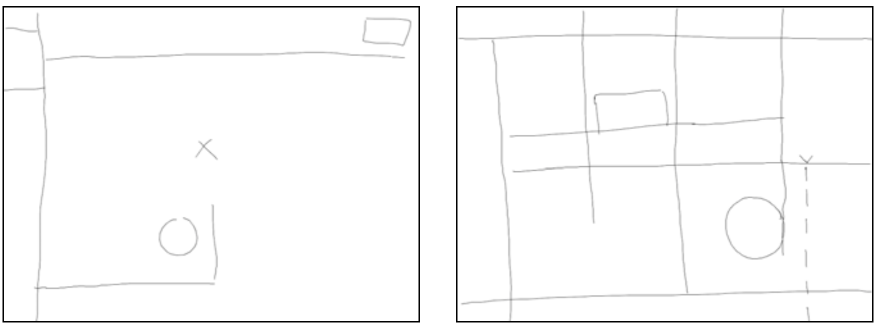


Fig. 2. The sketch to the left shows some route map characteristics, the one to the right is very similar to the stimulus (see Fig. 1, left) and shows survey characteristics, for example, there are many streets that are not used for direct routes between landmarks

Participants' subjective ratings were also analyzed. They were asked to order the three guide types by their suitability for the two tasks – searching for the YAH symbol and exploring the tactile map – by considering these questions: (1) how much did the particular guide type help you find the YAH symbol, and (2) how much did it hinder the exploration after the YAH symbol was found? The average ranks can be found in Table 2.

Table 2. Participants ranked the guide types when used in different tasks (1: best to 3: worst)

Task	**Searching the YAH Point**			**Exploring the map**		
Condition	Guiding line	Frame Marks	Grid	Guiding line	Frame Marks	Grid
Average Rank	1.50	1.25	3.00	2.00	1.00	3.00

The frame marks were ranked best for both tasks, the guiding line second and the grid were rated lowest. This is partially consistent with the quantitative findings (where the Grid condition was always worst). Notably, people can objectively find the YAH symbol faster with the Guiding lines than the Frame Marks, see Table 1, but they subjectively rank it lower.

During the experiments participants occasionally reported tactile illusions, primarily shortening of lines when comparing horizontal and vertical lines at a T-junction shape. That length-distorting effect resembles the Müller-Lyer illusion in vision and was reported before for congenitally blind people (cf. [26]).

A final note: The structure of the tracks was not completely identical in all maps. In the map with the frame marks one line segment had been inadvertently omitted. This resulted in one missing region and a change of +1 in the cardinality of one region and might have influenced the topological complexity of the map. But the missing segment was in the non-central part of the map that was not frequently explored and had no potential to be included in a route. It is unlikely the omitted segment had a significant impact on the results.

2.7 Discussion

Why were sighted participants run in this experiment, instead of blind people? It is a common belief that visually blind people have to perform better than sighted when asked to solve tasks on the basis of touch. Nevertheless, on the sensory level, blind and sighted people depend on the same bodily sensor system. In objective measures the two populations do not always differ. For example, there was no difference found for judgments of smoothness by sighted and blind participants, nor for passive touch (in which a stimulus is applied to a static skin surface), or for active touch (in which a stimulus is engaged with by a moving skin surface) [27]. Those findings indicate that there are primarily preference differences for the use of touch in blind and sighted people [28]. When it comes to basic cognitive tasks like object matching, which includes for example shape processing and recognition, visually impaired people outperform sighted people in the initial trials but after a few trials there is no difference [29]. It seems that haptic experience increases the speed of identification of abstract, simple shapes, but blind people do not perform as well when active elaboration is required [30]. In contrast to sensing and object matching, the experiment reported here depends on higher cognitive abilities that demand tacit and procedural knowledge of how to interpret a map. Therefore sighted people instead of visually impaired people were our participants. A future experiment will test visually impaired participants.

A closer analysis of the ratings of the sketches reveals that the sketches graded with 4 or worse were the only ones that showed route orientated encoding schema. In those sketches, only routes or part of a route were drawn, no other structures like streets around the area or that go from one end of the map to the other were displayed (see Figure 2, left). This was true of all the grade 5 and 6 sketches and over one third of the grade 4 sketches. It is possible that there was too much clutter in the map that masked the important parts (the distinction between clutter and relevant parts might vary depending on the task given). Clutter puts high cognitive load on the reader. Map-readers might employ strategies to focus on the part of the map with a lot of semantic content, i.e. the area where landmarks are located and the paths between them. Another interpretation could be that route concepts were acquired during exploration and only those could be externalized.

3 Proposal for Verbally Annotated Tactile Maps

3.1 Audio-Tactile Maps

One might assume that some form of compensation is needed in tactile maps because of their limited information. Compensation could be necessary for at least two problems. First, the tactile sense in comparison to the visual sense is relatively sparse. Only a limited amount of content can be displayed as it clutters quickly. One option could be to convey complementary content through another sensory channel. Content could be presented that was not encoded in the map due to space constraints. Another option would be to redundantly present content. Second, the tactile features of the map have to be explored serially, demanding high attention over a reasonably long period of exploration. Solutions that reduce the need to explore wide areas with the fingers could lower the necessary attention and manual effort.

Some of these ideas have been tested. Previous work investigated the use of sounds to provide additional content. Parkes and Lötzsch first presented audio-tactile devices to navigate and access non-textual, spatially distributed content [31, 32]. During the last two decades studies have considered computer driven audio-tactile devices, for example [33, 34], and computer-driven virtual audio-tactile maps, i.e. maps that are equipped with sound, for example, [35, 36] and verbal labels [37, 38]. Another approach was a force-feedback mouse and auditory labels to give directions in a mixed modal interface [39]. The multimodal approach turned out to be more comprehensive than the unimodal approach [40]. In the NOMAD system [32] a touch-sensitive pad provided raised features (as an analogue to map features) and verbal descriptions of the location that was touched. The descriptions contained information about what the user was touching and where this point was. Other work described the integration of verbal descriptions into tactile maps as well [33, 41, 42].

3.2 Verbally Annotated Tactile Maps

VAT maps are external multimodal representations that should have the potential to convey meaning that is stored in different types of mental representations. During an exploration of a VAT map, perceptual input from the tactile map that is believed to contribute to some form of mental spatial representation (analogous to the represented environment) is augmented with linguistic input (that is believed to be represented propositionally).

The underlying assumption for proposing Verbally Annotated Tactile maps is that different formats of mental representations can facilitate spatial reasoning. Research about learning with geographic maps that are presented in conjunction with a related text generally support this view (see [43] for a review). The cross-modal interaction of mental representations, i.e. non-propositionally encoded and propositionally encoded, should contribute to a common mental representation of the represented environment. Through the activation of concepts from different representational formats it might be possible to achieve an integration that contributes to the better understanding or faster access to what was learned from a VAT map. Given the growing evidence for functionally equivalent behavior on different inputs, such integration is likely, independent of perceptual or linguistic encoding (see [44] for a review).

The research literature suggests that combining propositional and spatial representation can benefit from *cross-modal effects*[2]. Few researchers have investigated the interaction of propositional representation and spatial representation when exploring a Verbally Annotated Tactile map. In such maps that the activation of concepts encoded in a tactile map may be facilitated by concepts encoded in the verbal description accompanying that map. Conveying map characteristics through concepts that were activated when listening to the verbal content may help the map-reader to better understand what is displayed in the map and how entities relate to each other. Conveying spatial and propositional representations, VAT maps might be an aid for visually impaired people navigating the world.

3.3 The Verbal Annotation System

Some mechanism to connect a tactile map with its verbal annotations is required. Some guiding ideas and questions concerning conceptual design are presented here.

The verbal annotations system could be connected to the tactile map through references that link certain point-of-interest (POI) or areas-of-interest (AOI) on the tactile map with some entity in the verbal annotation system. If the map-reader explores the vicinity of a POI or AOI, the verbal description should be presented, perhaps customized for approaching, passing or leaving. Calculating AOIs, POIs and descriptions could occur on the fly from a (dynamic) database that provides the map content. For descriptions, the verbal annotation system would contain a set of propositions and it should contain a rule system to determine which piece of content is to be presented in what way with the tactile map and the motion the map-reader executes. More precisely, to conceptually realize VAT maps, two aspects have to be considered. Each aspect relates to one distinct part of the verbal annotation system. The first one concerns the content of the propositions; the second one concerns the process of knowledge acquisition by the map-reader and the best way to support combining verbal content with a tactile map.

- "Which": Which content should be conveyed verbally that will not exceed the map-reader's cognitive capacities? Which content is selected to be verbalized to ensure a significant increase in the effectiveness of such maps for the map-reader? With some user modeling incorporated in the system, one could detect and individually support the map-user's exploration behavior.
- "How": How is the content best brought to the map-reader (for example at what point in time, for how much time, depending on what parameters)? How should the content to be structured with respect to the salience of objects in the environment; what should come first, what next (in terms of linearization of the information about space) etc.? Different strategies could be chosen, for example, fostering

[2] In this work and its context the concept does not mean learning in one modality and testing in another. Instead, human activities such as learning are supported by a cognitive system that is supposed to make use of spatial and propositional mental representations (see, e.g., the cognitive architecture proposed by Baddeley [45]).

a local perspective (by naming proximate landmarks before distal ones) vs. fostering global perspective, or focusing on the structure of the track network (by naming 2d landmarks, like intersections) vs. focusing on the structure of the built environment.

The verbal annotation system would also benefit from better mental models by

- Providing relevant redundant map content (as realized in some of the previous works reviewed above), or
- Providing relevant complementary map content, or
- Guiding the exploration strategies of the map-reader.

For providing redundant content, some knowledge captured in the map should be verbalized. For example, providing information about the intersection where the currently explored street ends. In this way, the map-reader does not need to extract the information from the map by exploring it in detail. Instead, she gets that information via spatial language sooner. Verbal complementary content could update a mental model with details that are not captured in the map. As there is lack of space in tactile maps, this content might add important details to be incorporated in planning or decision-making. To influence exploration, some survey information should be given informing the user about the local region, for example, which part of the map is she currently exploring and what other parts are located in which direction. In this way the exploration strategies that afford conceptualizing the content of the map could be supported. For example, when reaching an intersection (which is felt tactually) the user might hear an announcement about the points of interest that are in the vicinity (local perspective) and where the departing streets lead to (global perspective). Compared to a tactile map, a VAT map could be a type of map brings an advantage to visually impaired people, supporting them in gaining survey knowledge of the world for successful navigation.

3.4 Realizing Verbally Annotated Tactile Maps

In a VAT map the tactile map and the verbal annotations are independent components. The tactile map could be simulated or real: A simulated tactile map would be virtually explored with an actuator, for example, a force-feedback device that allows experiencing real forces when hitting a virtual object. Such a virtual map has the advantages that the interaction is very flexible and new environments can be simply loaded as new models. Unfortunately, 3d-models of particular environments, is generally not freely available. On the other hand, it is fairly easy to build such artificial environments. Tracking with such computer-controlled devices is also easy, as the actuator provides sensors to check the device's position in space. However, the spatial range of such devices is currently limited, usually smaller than a realistic tactile map. Additionally, there is, to my knowledge, no force-feedback interface that allows for multi-touch interaction or that lets the user employ kinaesthetic receptors to, for examples, feel deformations (with such a device there is only rigid, point-like touch). Furthermore, such devices are very expensive. If the tactile map were physical, real touch would be possible. But, to allow for dynamic verbalization some tracking technology must be in place (computer vision could be an option). Many of the drawbacks of virtual maps are avoided, but some advantages would be lost as well, for example, the integrated tracking. For real tactile maps, the technology to print them is affordable for small organizations or families. The verbal annotations to a tactile map could be realized through synthetic speech (the quality of such speech has improved a lot over the last few years), through dynamic assembly of pre-recorded audio snippets (this offers the best quality but is not flexible), or for experimental purposes by an experimenter or in a Wizard-of-Oz setting. Timely coordination and synchronization of both components is an important issue in all scenarios.

3.5 Future Research with Verbally Annotated Tactile Maps

The experiment with tactile maps that was reported here serves as one kind of baseline for future work developing an understanding of the effects of cross-modal interaction and how they can be employed to ease cognition for humans, particularly for those with visual-impairment. The basic design will be repeated with late-blind people, because there is good chance that they are able to perceive tactile pictures [46] and the organization of spatial processes seem to depend on the nature of perceptual input [7]. There is a good chance that they will know how to read a map. Knowing their strategies could inform work on how to teach other visually impaired persons to read and understand tactile maps [8].

Although there has been extensive work on visual-impairment (see Sections 1 & 2), there is only limited work that investigates cognition in blind people in relation to tactile maps. There are many open questions. For example, what is the maximum complexity of an environment that can be conveyed? Answering questions in this area would open the door to customized multimodal representations, for example VAT maps that make deliberate use of different cognitive subsystems to convey (spatial) meaning. An immediate critical question is, do visually impaired people have an advantage when they engage in wayfinding tasks after having consulted a VAT map.

To develop VAT maps there need to be some initial guidelines for which content is selected to be represented in the verbal annotation system. It is likely that the combination of verbal presentation in conjunction with a tactile map will have an influence on the exploration behavior of the map-reader. It is likely timing, granularity of presented content and context (for example task, knowledge about map conventions and others) will all influence the effectiveness of such systems. Further investigation is needed in what ways the map-reader is influenced when presented with verbal material about (1) the direct surroundings, so that she is accommodated in her current situation, (2) the adjacent areas, so that she has some knowledge of the areas she might wander into soon, and (3) the survey perspective that makes more distant landmarks accessible. We hypothesize that people who are presented with many pieces of "local content" would be more likely to build a mental model from connected representations of local phenomena. Such a mental model would hold many local details but probably not much survey knowledge. As result, tasks that build on survey knowledge should take longer to accomplish, other cognitive tasks that need detailed knowledge should be solved faster. Conversely, people who are presented with many pieces of "survey content" would be more likely build a survey like representation and will be good in querying it, but have trouble with questions regarding local details. Experiments to investigate these questions (and others) will tell us more about how a verbal annotation system should be constructed to accommodate different types of uses or contexts.

By measuring map exploration times, acquired spatial knowledge, and subjective measures, such as user satisfaction, the appropriateness of VAT maps to be used for spatial tasks could be evaluated and could be compared to tactile maps. If people have an advantage interacting with such maps then this would be an argument for trying to facilitate navigation with VAT maps. Comparison between tactile maps and VAT maps is needed.

4 Conclusion

The literature indicates that tactile maps are a valid option for conveying spatial knowledge non-visually. Comprehension and spatial reasoning was tested with Tactile You-Are-Here maps. The results showed that tactile YAH maps were understandable and map-readers performed well in spatial reasoning tasks. Different guiding types of YAH-symbol were tested. The less the guides interfered with the map content, the more useful they were. Map entities that were only useful for finding the YAH symbol but not for exploring and learning the map were judged to hinder exploration. In some cases they influenced learning in such a way that no survey model of the environment could be externalized (in sketch drawing). Having too many entities in tactile YAH maps likely hinder understanding the map.

To overcome this limit, some researchers favor a multimodal approach to conveying spatial knowledge to visually impaired people. Even if there is some evidence that in particular experiments there was no interaction between tactile and verbal channel [47], others have shown that reading tactile maps is improved by combining auditory and haptic information [41]. Going beyond the approaches that employ different *sensory modalities* this paper suggests an approach to employing different *representational modalities*. As an external multimodal representation I introduced the concept of a *Verbally Annotated Tactile map*. With this type of multimodal map, it might be possible to circumvent the bottleneck of low-detail and hard to interpret maps. Using VAT maps, inefficient navigation strategies might be avoided as people could build a richer model of what surrounds them through cross-modal interaction of propositional and spatial internal representations.

Issues that need to be addressed to fully realize the potential of VAT maps were discussed. The central issues to be addressed relate to the questions about which granularity of content best support navigation using a VAT map, which content should be verbalized or be represented tactually in a VAT map, and how should this content be structured and synchronized to map exploration behavior to best support the activities and strategies of the map-readers. Research in these directions has only just begun [48]. Through the investigation of the usage of VAT maps in spatial cognition tasks, and consideration of the value of such maps for visually impaired people when they engage in wayfinding tasks we may realize the goal of identifying the principles to successful construction of cognitively adequate tactile maps for navigation.

Acknowledgements. The research reported in this paper has been supported by the German Science Foundation (DFG) in the International Training Research Group 1247 "Cross-modal Interaction in Natural and Artificial Cognitive Systems" (CINACS). I thank all participants of the experiments, Prof. Habel and Kris Lohmann for past discussions, Franziska Dowall for the data collection and data processing, Doreen Dörfer and Niels Beuck for proof-reading, Pat McCrae and Thomas Shipley for their help with style and language, the Cognitive Systems Group at University of Bremen for the support and the anonymous reviewers for their helpful comments.

References

1. Kosslyn, S.M.: Image and mind. Harvard University Press, Cambridge (1980)
2. Weston, L., Handy, S.: Mental Maps. In: Hensher, D.A., Button, K.J., Haynes, K.E., Stopher, P. (eds.) Handbook of transport geography and spatial systems, pp. 533–546. Elsevier Science Ltd, Oxford (2004)
3. Tversky, B.: Spatial mental models. In: Bower, G.H. (ed.) The psychology of learning and motivation: Advances in research and theory, vol. 27, pp. 109–145. Academic Press Inc., San Diego (1991)
4. Forbus, K.D., Usher, J.: Sketching for knowledge capture: a progress report. In: Proceedings of the 7th International Conference on Intelligent User Interfaces, pp. 71–77. ACM, San Francisco (2002)
5. Freksa, C.: Qualitative Spatial Reasoning. In: Mark, D.M., Frank, A.U. (eds.) Cognitive and linguistic aspects of geographic space, pp. 361–372. Kluwer, Dordrecht (1991)
6. Espinosa, M.A., Ungar, S., Ochaíta, E., Blades, M., Spencer, C.: Comparing Methods for Introducing Blind and Visually Impaired People to Unfamiliar Urban Environments. Journal of Environmental Psychology 18, 277–287 (1998)
7. Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., Pietrini, P.: Imagery and spatial processes in blindness and visual impairment. Neuroscience & Biobehavioral Reviews 32, 1346–1360 (2008)
8. Thinus-Blanc, C., Gaunet, F.: Representation of Space in Blind Persons: Vision as a Spatial Sense? Psychological Bulletin 121, 20–42 (1997)
9. Gardiner, A., Perkins, C.: Here is the Beech tree! Understanding Tactile Maps in the Field. The Cartographic Journal 40, 277–282 (2003)
10. Levine, M.: You-Are-Here Maps: Psychological Considerations. Environment and Behavior 14, 221–237 (1982)
11. Richter, K.F., Klippel, A.: You-are-here maps: Wayfinding support as location based service. In: Moltgen, A., Wytzisk, A. (eds.) GI-Technologien für Verkehr und Logistik. Beiträge zu den Münsteraner GI-Tagen, pp. 363–382. IfGI Prints, Münster (2002)
12. Ungar, S.: Cognitive mapping without visual experience. In: Kitchin, R., Freundschuh, S. (eds.) Cognitive Mapping: Past, Present and Future, pp. 221–248. Routledge, London (2000)
13. Burnett, G., Lee, K.: The effect of vehicle navigation systems on the formation of cognitive maps. In: Underwood, G. (ed.) Traffic and Transport Psychology - Theory and Application, pp. 407–418. Elsevier, Amsterdam (2005)
14. Graf, C.: Kognitive Komplexität der Nutzung von Repräsentation räumlicher Umgebungen. Submitted to KogWis 2010 (2010)
15. Habel, C., Graf, C.: Towards Audio-tactile You-are-here Maps: Navigation Aids for Visually Impaired People. In: Workshop Proceedings You-Are-Here-Maps. SFB/TR8 Spatial Cognition and University of Freiburg, Freiburg (Breisgau), Germany, pp. 1–10 (2008)
16. Richter, A.Z.: Using focus maps to ease map reading: developing smart applications for mobile devices. Kunstliche Intelligenz 4 (2002)

17. Rice, M., Jacobson, R.D., Golledge, R.G., Jones, D.: Design Considerations for Haptic and Auditory Map Interfaces. Cartography and Geographic Information Science 32, 381–391 (2005)
18. Pun, T., Roth, P., Bologna, G., Moustakas, K., Tzovaras, D.: Image and video processing for visually handicapped people. Journal on Image Video Processing 2007, 1–12 (2007)
19. Castner, H.W.: Tactual Maps And Graphics - Some Implications For Our Study of Visual Cartographic Communication. Cartographica: The International Journal for Geographic Information and Geovisualization 20, 1–16 (1983)
20. Gardiner, A., Perkins, C.: 'It's a sort of echo...': Sensory perception of the environment as an aid to tactile map design. British Journal of Visual Impairment 23, 84–91 (2005)
21. Lederman, S.J., Browse, R.A.: The physiology and psychophysics of touch. In: Dario, P. (ed.) Proceedings of a NATO Advanced Research Workshop on Sensors and sensory systems for advanced robots, pp. 71–91. Springer, Berlin (1988)
22. Jehoel, S., McCallum, D., Rowell, J., Ungar, S.: An empirical approach on the design of tactile maps and diagrams: The cognitive tactualization approach. British Journal of Visual Impairment 24, 67–75 (2006)
23. Michel, R.: Interaktiver Layoutentwurf für individuelle taktile Karten. Shaker, Aachen (2000)
24. Ungar, S., Blades, M., Spencer, C.: Strategies for knowledge acquisition from cartographic maps by blind and visually impaired adults. Cartographic Journal 34, 93–110 (1997)
25. Gardner, J.A., Bulatov, V.: Directly Accessible Mainstream Graphical Information. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 739–744. Springer, Heidelberg (2004)
26. Millar, S., Al-Attar, Z.: Illusions in reading maps by touch: Reducing distance errors. British Journal of Psychology 92, 643–657 (2001)
27. Heller, M.A.: Texture perception in sighted and blind observers. Perception & Psychophysics 45, 49–54 (1989)
28. Jehoel, S., Ungar, S., McCallum, D., Rowell, J.: An Evaluation of Substrates for Tactile Maps and Diagrams: Scanning Speed and Users' Preferences. Journal of Visual Impairment and Blindness 99, 85–95 (2005)
29. Postma, A., Zuidhoek, S., Noordzij, M.L., Kappers, A.M.: Differences between early blind, late blind and blindfolded sighted people in haptic spatial configuration learning and resulting memory traces. Perception 36, 1253–1265 (2007)
30. Vecchi, T.: Visuo-spatial imagery in congenitally totally blind people. Memory 6, 91–102 (1998)
31. Lötzsch, J.: Computer-aided access to tactile graphics for the blind. In: Zagler, W.L., Busby, G., Wagner, R.R. (eds.) ICCHP 1994. LNCS, vol. 860, pp. 575–581. Springer, Heidelberg (1994)
32. Parkes, D.: NOMAD: An audio-tactile tool for the acquisition, use and management of spatially distributed information by visually impaired people. In: Proceedings of the second international symposium on maps and graphics for visually disabled people, pp. 24–29. Nottingham University, Nottingham (1988)
33. Parente, P., Bishop, G.: BATS: the blind audio tactile mapping system. In: Proceddings of the ACM Southeast Regional Conference. ACM, Savannah (2003)
34. Wall, S., Brewster, S.: Feeling what you hear: tactile feedback for navigation of audio graphs. In: Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 1123–1132. ACM, Montréal (2006)

35. Nasir, T., Roberts, J.C.: Sonification of Spatial Data. In: The 13th International Conference on Auditory Display (ICAD 2007), Schulich School of Music, McGill University, Canada, pp. 112–119 (2007)
36. Zhao, H., Plaisant, C., Shneiderman, B.: iSonic: interactive sonification for non-visual data exploration. In: Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 194–195. ACM, New York (2005)
37. Jacobson, R.D.: Navigating maps with little or no sight: An audio-tactile approach. In: Proceedings of the Workshop on Content Visualization and Intermedia Representations (CVIR 1998), Montréal, Québec, Canada (1998)
38. Miele, J.A., Landau, S., Gilden, D.: Talking TMAP: Automated generation of audio-tactile maps using Smith-Kettlewell's TMAP software. British Journal of Visual Impairment 24, 93 (2006)
39. Jacobson, R.D., Kitchin, R., Golledge, R.G.: Multimodal virtual reality for presenting geographic information. In: Unwin, D.J., Fisher, P. (eds.) Virtual Reality in Geography, pp. 382–400. CRC Press, London (2002)
40. Golledge, R.G., Rice, M., Jacobson, R.D.: A Commentary on the Use of Touch for Accessing On-Screen Spatial Representations: The Process of Experiencing Haptic Maps and Graphics. The Professional Geographer 57, 339–349 (2005)
41. Holmes, E., Jansson, G.: A Touch Tablet Enhanced with Synthetic Speech as a Display for Visually Impaired People's Reading of Virtual Maps. Presented at the 12th Annual International Technology and Persons with Disabilities conference (CSUN 1997), Los Angeles, CA, March 18-22 (1997)
42. Siekierska, E., Muller, A.: Tactile and Audio-Tactile Maps within the Canadian Government On-Line Program. The Cartographic Journal 40, 299–304 (2003)
43. Verdi, M., Kulhavy, R.: Learning with Maps and Texts: An Overview. Educational Psychology Review 14, 27–46 (2002)
44. Loomis, J.M., Klatzky, R.L., Avraamides, M., Lippa, Y., Golledge, R.G.: Functional Equivalence of Spatial Images Produced by Perception and Spatial Language. In: Mast, F., Jäncke, L. (eds.) Spatial Processing in Navigation, Imagery and Perception, pp. 29–48. Springer, New York (2007)
45. Baddeley, A.: Working memory: looking back and looking forward. Nature Reviews Neuroscience 4, 829–839 (2003)
46. Heller, M.A.: Picture and pattern perception in the sighted and the blind: the advantage of the late blind. Perception 18, 379–389 (1989)
47. Jansson, G., Juhasz, I., Cammilton, A.: Reading virtual maps with a haptic mouse: Effects of some modifications of the tactile and audio-tactile information. British Journal of Visual Impairment 24 (2006)
48. Habel, C., Kerzel, M., Lohmann, K.: Verbal Assistance in Tactile-Map Explorations: A Case for Visual Representations and Reasoning. In: Proceedings of AAAI 2010. AAAI Press / The MIT Press (2010) (to be published)

Generating Adaptive Route Instructions Using Hierarchical Reinforcement Learning

Heriberto Cuayáhuitl[1], Nina Dethlefs[2], Lutz Frommberger[1], Kai-Florian Richter[1], and John Bateman[1,2]

[1] Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition, University of Bremen
Enrique-Schmidt-Str. 5, 28359, Bremen, Germany
[2] FB10 Faculty of Linguistics and Literary Sciences, University of Bremen
Bibliothekstrasse 1, 28359, Bremen, Germany
heriberto@uni-bremen.de

Abstract. We present a learning approach for efficiently inducing adaptive behaviour of route instructions. For such a purpose we propose a two-stage approach to learn a hierarchy of wayfinding strategies using hierarchical reinforcement learning. Whilst the first stage learns low-level behaviour, the second stage focuses on learning high-level behaviour. In our proposed approach, only the latter is to be applied at runtime in user-machine interactions. Our experiments are based on an indoor navigation scenario for a building that is complex to navigate. We compared our approach with flat reinforcement learning and a fully-learnt hierarchical approach. Our experimental results show that our proposed approach learns significantly faster than the baseline approaches. In addition, the learnt behaviour shows to adapt to the type of user and structure of the spatial environment. This approach is attractive to automatic route giving since it combines fast learning with adaptive behaviour.

1 Introduction

Adaptive route instructions provide a human user with information that is specifically suitable to his or her needs by taking the current situation and individual properties of the user into account. Such kind of adaptation is achieved along two dimensions: (a) the cognitive principles underlying human route descriptions and (b) models of different user types who may differ in their informational needs. The first aspect is related to what Lovelace et al. [1] stated as characteristics of good route descriptions, which are based on experimental findings involving human route descriptions. These include the preference of landmarks over metrical distances or purely path-based information, the inclusion of confirmatory information on long route segments to prevent the user from going too far, and limiting the amount of detail to a necessary minimum to avoid redundancy. The second aspect is related to the fact that users may differ in their amount of prior knowledge and hence in their informational needs or preferences. For example,

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 319–334, 2010.

a user who is (partially) familiar with an environment might prefer the shortest route to some destination, while an unfamiliar user might prefer the easiest route in order not to get lost. Similarly, while the first type of user might be able to find their way with a limited amount of detail and perceive too much redundancy as disturbing or patronizing, the latter type of user will likely require more guidance.

In this paper, we address the topic of generating online adaptive route instructions using *reinforcement learning* (RL) [2]. In order to achieve adaptive route instructions while keeping the system efficient, as it is needed in navigation systems, for example, we propose to apply a two-stage approach to learn a hierarchy of wayfinding strategies employing *hierarchical reinforcement learning* (HRL). In a first stage, we learn low-level behaviour, such as 'left' or 'right.' In a second stage, we learn high-level behaviour, such as 'go to the next junction.' The latter behaviour can be applied in online user-machine interactions to provide online adaptation and interactive route guidance. This enables the system to identify a user's prior knowledge and information needs during an interaction and to flexibly adapt to them. We present experimental results, based on a simulated environment, that compare flat RL with HRL fully-online and semi-online behaviour. We demonstrate that the semi-online policies are learnt faster and more efficiently than the others and hence are suitable for online learning and adaptation.

This paper is organized as follows: Section 2 gives an overview of related work on adaptive route generation. Sections 3 and 4 introduce the approaches of reinforcement learning and hierarchical reinforcement learning, respectively, while the performance of fully-online and semi-online learnt behaviours are compared in Section 5. In that section, we will show that our proposed approach is promising for its application in online user-machine interaction and that it can be used for inducing adaptive behaviour in in-situ navigation assistance.

2 Related Work on Route Instruction Generation

Related work on generation of adapative route instructions has addressed several issues, including the usage of landmarks, cognitive adaquacy of route descriptions in changing spatial situations, and the tailoring of route descriptions towards different user groups. Landmarks are highly prominent features in route directions [3], both because they are crucial elements in structuring human knowledge about an environment [4] and because they are powerful means to link wayfinding actions to the points along a route where they have to be performed, i.e., in providing procedural information [1]. In general human direction givers try to provide concise instructions, in order not to overload the receiver with too much information [5]. The procedural nature of landmarks makes them essential elements in human route directions, which can also be attributed to cognitive landmark-based mechanisms of combining several consecutive instructions into a single, higher-order instruction, which is more compact (so called spatial chunking, [6]). Not surprisingly, landmarks were identified as one key feature of improving automated navigation services [7, 8]. However, today's commercially available navigation services hardly include any landmarks at all. They produce instructions in a uniform, repetitive form, referring to street names and metric distances only.

In addition to the inclusion of landmarks in route directions, a further issue of major relevance in the context of this paper is adaptation to changing spatial situations. As pointed out by Klippel et al. [9], people vary descriptions on what to do at intersections to both what action has to be performed (a functional aspect) and to the layout of the intersection, including the presence of landmarks (a structural aspect). More generally, people adapt instructions to the assumed recipients' information needs, which includes taking into account the transportation means and the required detail (scale) to locate the destination (e.g., an ATM mounted on a wall of a bank compared to the bank building itself) [10]. Attempts to capture this need for adaptation in automated services were pursued on different levels. Some approaches aimed for adapting the underlying route to travel between origin and destination by searching for the simplest path [11], the most reliable path, i.e., the one with the least probability for wayfinding errors [12], or the path that is the simplest to describe [13].

Systems that have attempted to integrate several of the abovementioned principles include Dale et al.'s CORAL system [14]. It uses data that is available from existing geographical databases to identify relevant information for describing a route, which includes features that may serve as landmarks. The authors apply natural language generation methods to construct instructions that adapt to the given environment. Namely, these methods are segmentation (comparable to spatial chunking), which groups single instructions into higher-level instructions, and aggregation, which is a process to construct sentences that communicate several pieces of information at once. A similar approach was presented by Richter [15]. His work focused on the steps before the actual information presentation, i.e., the identification of relevant information and spatial chunking. It includes methods for determining the functional roles of different kinds of landmark candidates, among them point-like, linear, area-like features, as well as intersection structure (cf. [16]), and an optimization process to find the 'best' combination of chunks (e.g., the most concise description of a route).

Research that applied machine learning to route generation includes work by Marciniak and Strube [17, 18], who used machine learning techniques to generate natural language route directions. Using a manually annotated corpus of existing route directions as input, they had their system learn K* classifiers [19] for individual elements of a vector representation of an expression's form. Overall expression generation is then realized as a cascade of these classifiers, where later classifiers can make use of the output of earlier classifiers. Expressions are represented as tree adjoining grammars (TAGs).

We are not aware of any prior work that has applied reinforcement learning to adaptive route instruction generation. In the rest of the paper we show that hierarchical reinforcement learning (based on the divide-and-conquer approach) is a promising method to optimize—in an efficient and scalable way—the behaviour of adaptive route instruction generation systems.

3 Reinforcement Learning for Route Instruction Generation

3.1 Adapting to a User's Behaviour with Reinforcement Learning

Reinforcement learning is particularily useful if the characteristics of an underlying system are unknown or only partially known. This is what we encounter when generating adaptive route instructions. Generated route instructions so far make assumptions on how a user can cope with the produced directions. While these assumptions may hold for a majority of users, it does not take special skills or preferences of an individuum into account. For example, a user familiar with an environment will prefer different instructions than someone visiting this place for the first time. Presented with instructions aiming at the experts' knowledge, the unexperienced user will most likely fail to follow the instructions correctly. Other users may just have shifted preferences (like counting crossroads or looking for landmarks). These preferences can be assumed, but usually not fully described, and are a matter of observation of the user's interaction with the environment.

This observed interaction is the fundament of reinforcement learning, as the learned behaviour (the *policy*) is adapted by rewards depending on actions agents perform within their environment. RL is able to produce an optimal policy for a system that cannot be described in a closed form simply by interacting with it. Thus, it becomes a valuable means for generating adaptive route instructions.

3.2 Formalization of the Learning Task

A user-machine interaction for route instruction generation can be described as follows: The machine receives a user input (e.g., a query requesting instructions or the user's current location) to navigate from an origin to a destination. It enters such information into its knowledge base and then extracts an environment state s to choose an instruction (also referred to as 'action' a). Then, the chosen action is used to generate the corresponding output conveyed to the user. These kinds of interactions are followed in an iterative process until the goal is reached. Assuming that the machine receives a numerical reward r for executing action a when the spatial environment makes a transition from state s_t (at time t) to a next state s_{t+1}, an interaction can be expressed as $\{s_1, a_1, r_2, s_2, a_2, r_3, ..., s_{T-1}, a_{T-1}, r_T, s_T\}$, where T is the final time step.

Such sequences can be used by a reinforcement learning agent to optimize the behaviour of the route instruction controller. A reinforcement learning wayfinding agent aims to learn its behaviour from interaction with an environment, where situations are mapped to actions by maximizing a long-term reward signal. The standard reinforcement learning paradigm makes use of the formalism of Markov Decision Processes (MDPs) [2]. An MDP is characterized by a set of states S, a set of actions A, a state transition function T, and a reward or performance function R that rewards the agent for each selected action. A given sequence of states, actions, and rewards receives a total cumulative discounted reward expressed as

$$r = r_1 + \gamma r_2 + \gamma^2 r_3 + ... \gamma^{\tau-1} r_\tau = \sum_{k=0}^{\tau-1} \gamma^k r_{k+1}, \tag{1}$$

where the discount rate $0 \leq \gamma \leq 1$ makes future rewards less valuable than immediate rewards as it approaches 0. In the equation above, the term on the right-hand side is referred to as 'the expected value of the reward,' and can be computed recursively using a state-value function $V^\pi(s)$, which returns the value of starting in state s and then following policy π thereafter. The value-function is defined by the Bellman equation for V^π expressed as

$$V^\pi(s) = \sum_a \pi(s,a) \sum_{s'} P(s'|s,a) \left[R(s'|s,a) + \gamma V^\pi(s') \right]. \tag{2}$$

Alternatively, the expected value of the reward can also be computed recursively using an action-value function $Q^\pi(s,a)$, which returns the cumulative reward of starting in state s, taking action a and then following policy π thereafter. The action-value function is defined by the Bellman equation for Q^π expressed as

$$Q^\pi(s,a) = \sum_{s'} P(s'|s,a) \left[R(s'|s,a) + \gamma V^\pi(s') \right]. \tag{3}$$

An optimal policy π^* can be found by using the following Bellman equations that represent a system of equations, one for each state:

$$V^*(s) = \max_a \sum_{s'} P(s'|s,a) \left[R(s'|s,a) + \gamma V^*(s') \right], \tag{4}$$

or state-action pair:

$$Q^*(s,a) = \sum_{s'} P(s'|s,a) \left[R(s'|s,a) + \gamma \max_{a'} Q^*(s',a') \right]. \tag{5}$$

Finally, an optimal policy performs action selection according to

$$\pi^*(s) = \arg\max_{a \in A} Q^*(s,a). \tag{6}$$

The optimal policy π^* can be found using reinforcement learning algorithms such as Q-learning or SARSA [2]. However, the flat tabular reinforcement learning framework lacks scalability, that is, it requires very long training times as the problems become more complex and learning becomes infeasible for anything but very small state spaces [20]. But for generating online route instructions we must rely on fast training times and immediate adaptations. Thus, we use a hierarchical reinforcement learning framework that decomposes an MDP into a hierarchy of Semi-Markov decision processes (each one representing a subsequence of instructions). This approach has been successfully applied to spoken dialogue agents with large state spaces [21, 22].

4 Hierarchical Reinforcement Learning Approaches for Route Instruction Generation

In this research we treat the task of route instruction generation as a discrete Semi-Markov Decision Process (SMDP) in order to address the problem of scalable optimization. A discrete-time SMDP $M = <S, A, T, R>$ is characterized by a set of states S; a set of actions A; a transition function T that specifies the next state s' given the current state s and action a with probability $P(s', \tau|s, a)$; and a reward function $R(s', \tau|s, a)$ that specifies the reward given to the agent for choosing action a when the environment makes a transition from state s to state s'. The random variable τ denotes the number of time-steps taken to execute action a in state s. This formulation allows temporal abstraction, where actions take a variable amount of time to complete their execution [23]. In this model two types of actions can be distinguished: (a) primitive actions roughly corresponding to low-level instructions (e.g., straight, right, left, turn around), and (b) composite actions corresponding to sub-sequences of instructions (e.g., go to the other end of the corridor). Whilst the root and children models execute primitive and composite actions, the grandchildren models—at the bottom of the hierarchy—execute only primitive actions.

4.1 Fully-Online Hierarchical Learning

This research treats each composite action as a separate SMDP as described in Cuayáhuitl [21]. In this way an MDP can be decomposed into multiple SMDPs hierarchically organized into L levels and N models per level. We denote each SMDP as $\mathcal{M} = \{M_j^i\}$, where $j \in \{0, ..., N-1\}$ and $i \in \{0, ..., L-1\}$. Thus, any given SMDP in the hierarchy is denoted as $M_j^i = <S_j^i, A_j^i, T_j^i, R_j^i>$, see Figure 1 for an illustration. The indexes i and j only identify an SMDP in a unique way in the hierarchy, they do not specify the execution sequence of SMDPs because that is learnt by the reinforcement learning agent.

The goal in an SMDP is to find an optimal policy π^* that maximizes the reward of each visited state. The optimal action-value function $Q^*(s, a)$ specifies the expected cumulative reward for executing action a in s and then following π^*. The Bellman equations for Q^* of subtask M_j^i can be expressed as

$$Q_j^{*i}(s, a) = \sum_{s', \tau} P_j^i(s', \tau|s, a)[R_j^i(s', \tau|s, a) + \gamma^\tau \max_{a'} Q_j^{*i}(s', a')], \tag{7}$$

The optimal policy for each model in the hierarchy is defined by

$$\pi_j^{*i}(s) = \arg\max_{a \in A_j^i} Q_j^{*i}(s, a). \tag{8}$$

These policies can be found by reinforcement learning algorithms for SMDPs such as the Hierarchical Semi-Markov Q-Learning algorithm (also referred to as HSMQ-Learning) [24]. In this algorithm, the action-value function Q_j^{*i} of equation 7 is approximated according to

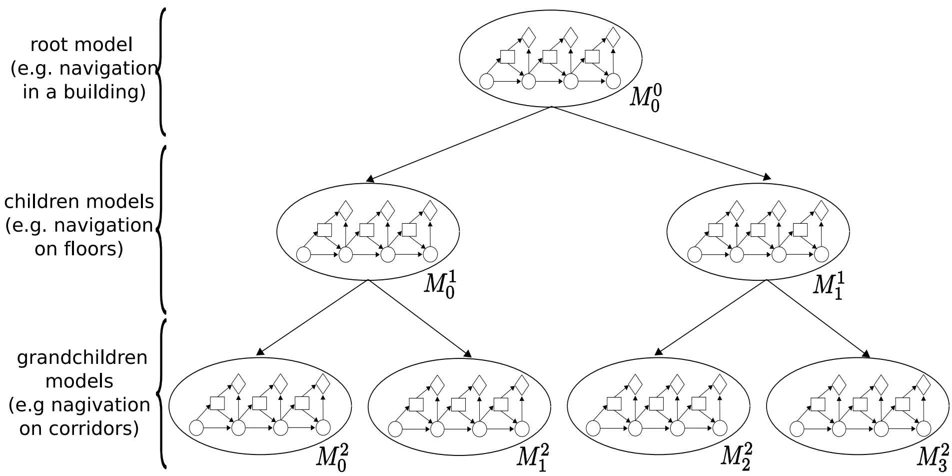


Fig. 1. Hierarchy of SMDPs M_j^i, where i denotes a level and j the model per level. In each SMDP, bottom circles represent environment states, rectangles represent actions, and diamonds represent rewards.

$$Q_j^i(s_t, a_t) \leftarrow (1-\alpha) Q_j^i(s_t, a_t) + \alpha \left[r_{t+\tau} + \gamma^\tau \max_{a'} Q_j^i(s_{t+\tau}, a') \right]. \quad (9)$$

The summation over all τ time steps as they appear in equation 7 is reflected here by using cumulative rewards $r_{t+\tau} = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + ... + \gamma^{\tau-1} r_{t+\tau}$ received for executing actions a_t, and by raising γ to the power τ.

4.2 Semi-online Hierarchical Learning

Because efficient learning is required at runtime (i.e., users should not wait too long for receiving route instructions), we propose a two-stage approach to speed up the hierarchical method above. The first stage—applied before user-machine interaction—induces the behaviour of the policies at the bottom of the hierarchy, expressed as

$$\pi_j^{*L-1}(s) = \arg \max_{a \in A_j^{L-1}} Q_j^{*L-1}(s, a), \quad (10)$$

where the index $L-1$ represents the bottom layer in the hierarchy. The second stage—applied at runtime user-machine interaction—induces the behaviour of parent agents and the bottom agent including the goal state, expressed as

$$\pi_j^{*i}(s) = \arg \max_{a \in A_j^i} Q_j^{*i}(s, a), \forall i \neq L-1 \vee s^g \in S_j^i, \quad (11)$$

where s^g is the goal state. Our approach avoids learning from scratch by reusing learnt behaviour from the first stage, and focuses on learning high-level behaviour. The same learning algorithm can also be used to find the policies above.

5 Experiments and Results

The aim of our experiments was: (1) to provide a proof-of-concept application of our semi-online learning approach by comparing it against fully-online flat and hierarchical learning, and (2) to induce adaptive behaviour for familiar and unfamiliar users. Our experiments are based on indoor route instruction generation, and use a simulated spatial environment with data derived from a real building that is complex to navigate. We employ data from a single floor, which is represented as an indirected graph (see Fig. 2). Such a graph and the stochastic behaviour described in subsection 5.2 form the agent's learning environment.

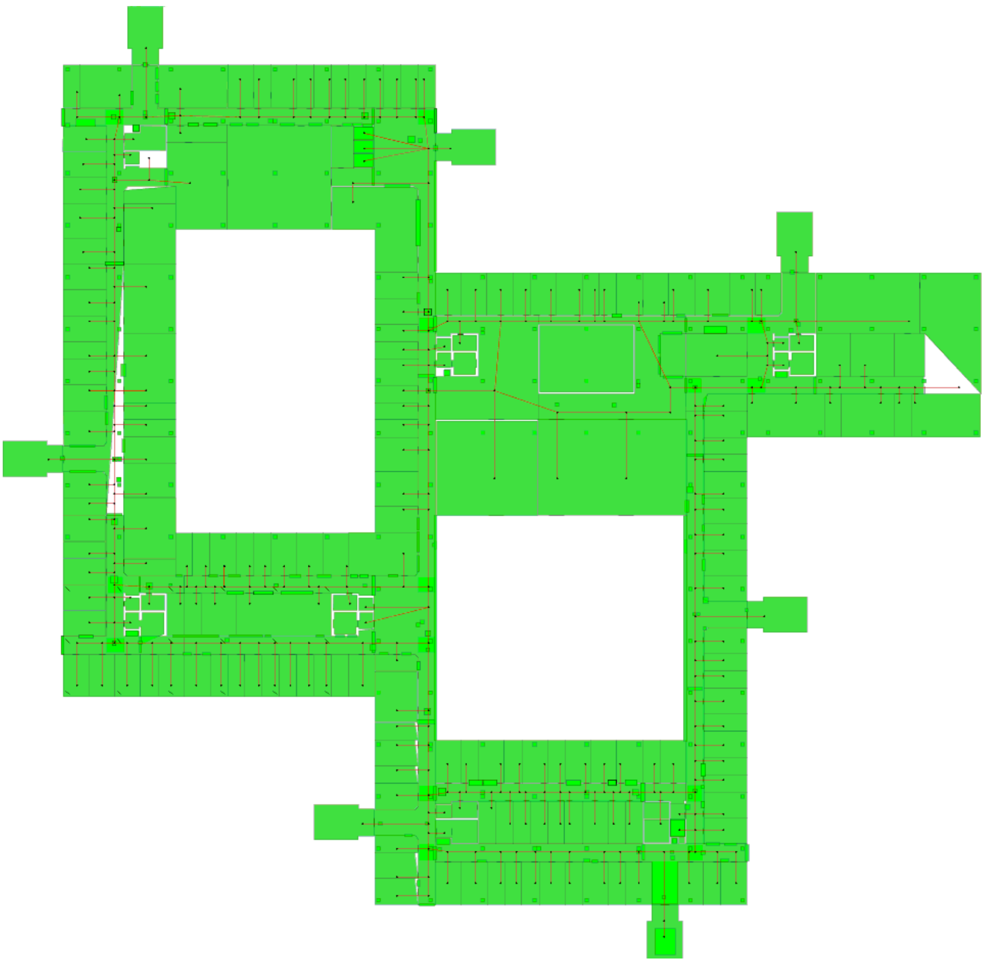


Fig. 2. Map showing an undirected graph of the spatial data used for learning navigational behaviour. The small black circles represent graph nodes, which are connected with the corresponding edges so that navigation is allowed between corridors and rooms.

5.1 Experimental Setting

Flat reinforcement learning used the state space representation shown in Table 1, and the action set shown in Table 2, resulting in a space of 896000 state-actions. Whilst the first two state variables (location and orientation) represent basic information that can be used for local navigational decisions, the other state variables (score of best landmark, segment length, segments to follow, and user type) represent information of the route trajectory that can be used for more informative navigational decisions. The latter can take into account the inclusion or exclusion of landmarks and using a trajectory that is less confusing to follow according to the given stochastic environment.

The reward function addresses compact instructions based on the shortest number of turns. It is defined by the following rewards given to the reinforcement learning agent for choosing action a when the environment makes a transition from state s to state s':

$$r(s, a, s') = \begin{cases} 0 \text{ for reaching the goal state} \\ \text{-1 for a straight action.} \\ \text{-10} \quad \text{otherwise.} \end{cases} \quad (12)$$

For hierarchical learning, the state-action space representation used 82 models (one parent and 81 children) induced automatically from turning points in the spatial environment. Whilst the root agent used child agents instead of the action

Table 1. State variables for generating indoor route instructions. Flat learning used a graph with 400 nodes, corresponding to $|S| = 400 * 4 * 5 * 2 * 4 * 2 = 128000$ states. We follow Raubal & Winter [25] to rank landmarks based on the categories suggested by Sorrows & Hirtle [4].

Variable	Values	Description
Location	(x,y)	Coordinates of graph nodes from the space
Orientation	$\{0, 1, 2, 3\}$	Corresponding to 'south', 'north', 'east', 'west'
Score of best landmark	$\{0, 1, 2, 3, 4\}$	The higher the more suitable
Segment length	$\{0, 1\}$	Corresponding to 'short' and 'long'
Segments to follow	$\{0, 1, 2, 3\}$	Number of segments in the location
User type	$\{0, 1\}$	Corresponding to 'unfamiliar' and 'familiar'

Table 2. Primitive actions for generating indoor route instructions

Action	Description
straight	go straight
left_nolandmark	turn left without referring to a present landmark
left_landmark	turn left referring to a present landmark
right_nolandmark	turn right without referring to a present landmark
right_landmark	turn right referring to a present landmark
turnaround_nolandmark	turn around without referring to a present landmark
turnaround_landmark	turn around referring to a present landmark

'straight', the child agents used the same action set as the flat learning agent. In addition, these experiments used the same state transition and reward functions as in flat learning, used in each agent.

The learning setup used Q-Learning for flat reinforcement learning [2] and HSMQ-Learning for hierarchical reinforcement learning (briefly described in the previous section). The learning parameters used by the algorithms were the same for all learning approaches. The learning rate parameter α decays from 1 to 0 according to $\alpha = \frac{100}{(100+\tau)}$, where τ represents elapsed time-steps in the current subtask. Each subtask M_j^i had its own learning rate. The discount factor $\gamma = 1$ makes future rewards as valuable as immediate rewards. The action selection strategy used ϵ-Greedy with $\epsilon = 0.01$, and initial Q-values of 0.

5.2 Stochastic Behaviour in the Learning Environment

The conditional probabilities shown in Table 3 were used to model user confusion in indoor wayfinding. Such probabilities were hand-crafted according to the following assumptions.

- Unfamiliar users get more confused than familiar users.
- Goal destinations within short segments (corridors) are easier to find than within long segments.
- Users get more confused at junctions with three or more segments than with two or less segments.
- The confusion is lower when following suitable landmarks (score ≥ 3), and higher when following less suitable landmarks (score ≤ 2).

Finally, because the probabilities above are location-independent for the spatial environment, we used additional location-dependent probabilities expressed as

$$Pr(UserConfusion|TurningPoint, Orientation, UserType) \quad (13)$$

for modelling user confusion, specifically at turning points.

5.3 Experimental Results

Figure 3 shows the learning curves of policies for an average-sized route instruction, averaged over 10 training runs of 10^4 episodes. It can be observed that our proposed approach learned faster than the other approaches. Whilst the latter two require four orders of magnitude (10^4 episodes), our proposed approach required only three orders of magnitude (10^3 episodes). The fact that our proposed approach stabilizes its behaviour much more quickly than the others suggests that it has promising application for online user-machine interaction. To give an idea of learning times, the instructions shown in Table 4 took 3.2 seconds to generate using 1000 episodes on a MacBook computer (processor Intel Core 2 Duo of 2.4 GHz and 2GB in RAM).

The first column of the sample set shown in Table 4 shows the navigation policies π_j^{*i} used by the hierarchical learning agent, where index j of the lower

Table 3. Conditional probability table to model user confusion in wayfinding. The first four columns are state variables (see Fig. 1 for a description of their values) and the last two columns are state transition probabilities. Notation: s = current environment state, $\tilde{s'}$ =distorted next state, a_{wl} =action with landmark, a_{wol} =action without landmark.

User Type	Segment Length	Segments to Follow	Landmark Weight	$P(\tilde{s'}\|s, a_{wl})$	$P(\tilde{s'}\|s, a_{wol})$
0	0	≤ 2	≤ 2	.10	.15
0	0	≤ 2	≥ 3	.10	.10
0	0	≥ 3	≤ 2	.10	.25
0	0	≥ 3	≥ 3	.10	.20
0	1	≤ 2	≤ 2	.10	.15
0	1	≤ 2	≥ 3	.10	.10
0	1	≥ 3	≤ 2	.10	.35
0	1	≥ 3	≥ 3	.10	.30
1	0	≤ 2	≤ 2	.05	.10
1	0	≤ 2	≥ 3	.05	.05
1	0	≥ 3	≤ 2	.05	.15
1	0	≥ 3	≥ 3	.05	.10
1	1	≤ 2	≤ 2	.05	.10
1	1	≤ 2	≥ 3	.05	.05
1	1	≥ 3	≤ 2	.05	.20
1	1	≥ 3	≥ 3	.05	.15

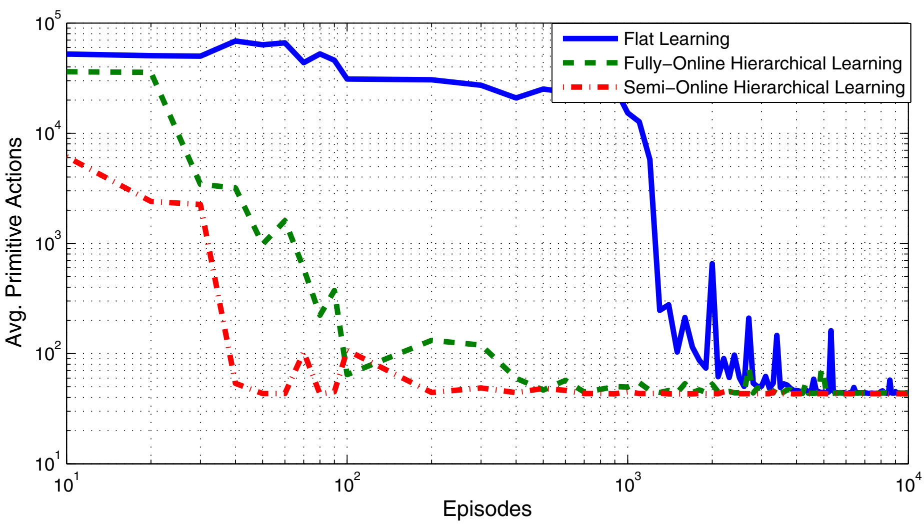


Fig. 3. Learning curves of route instruction strategies with three learning approaches. We plot the average number of primitive actions because our learning agents aim for efficient trajectories according to the given stochastic environment. Alternatively, other measures can be used, such as average number of textual route instructions.

Table 4. Sample generation of route instructions using a hierarchy of policies π_j^{*i}, where π_0^{*0} is the root policy and π_j^{*1} are the children policies. Whilst the former is used to navigate from origin to destination, the latter are used to navigate across segments. The term 'segment' used here denotes a route with a start and end junctions, but without intermediate junctions.

Policy	Environment State	Actions	Interpretation
π_0^{*0}	(2250,1330),3,0,0,0,0	$\pi_{307-314}^{*1}$	Go straight until
$\pi_{307-314}^{*1}$	(2250,1330),3,0,0,0,0	straight (3 times)	the second junction.
π_0^{*0}	(1690,1260),3,1,0,3,0	$\pi_{314-192}^{*1}$	
$\pi_{314-192}^{*1}$	(1690,1260),3,1,0,3,0	straight	
π_0^{*0}	(2250,1330),3,0,0,0,0	left_landmark	Turn left at <landmark>
π_0^{*0}	(1480,1260),0,1,1,3,0	$\pi_{192-1000}^{*1}$	and go straight until
$\pi_{192-1000}^{*1}$	(1480,1260),0,1,1,3,0	straight (six times)	the next junction.
π_0^{*0}	(1480,1885),0,2,1,3,0	right_landmark, $\pi_{1000-110}^{*1}$	Turn right at <landmark>
$\pi_{1000-110}^{*1}$	(1480,1885),3,3,1,3,0	straight (15 times)	and go straight until
			the next junction.
π_0^{*0}	(480,1880),3,1,1,3,0	right_landmark, π_{110-89}^{*1}	Turn right at <landmark>
π_{110-89}^{*1}	(480,1880),1,3,1,3,0	straight (9 times)	and go straight until
π_0^{*0}	(480,1480),1,1,1,3,0	π_{89-15}^{*0}	<landmark>, the destination
π_{89-15}^{*0}	(480,1480),1,1,1,3,0	straight (4 times)	will be at your right.
		right_landmark, straight	
π_0^{*0}	(580,1260),2,1,1,0,0	[goal state]	

level of the hierarchy represents a route segment with graph nodes 'start-end'. For example, policy $\pi_{192-1000}^{*1}$ navigates from node 192 to node 1000. The second column shows the environment state using the state variables described in Table 1. The third column shows the actions taken by the agent; recall that whilst the hierarchical learning agents take primitive actions a and composite actions π_j^{*i}, the flat learning agent only takes primitive actions. The fourth and final column provides an interpretation of chunked route instructions, where the chunking is based on major changes of direction such as left or right.

Our preliminary results show that the learnt policies include more landmarks for unfamiliar users, which would result in longer natural language instructions that provide a better anchoring of actions in the environment. In addition, we observed that the learnt behaviour can provide tailored instructions for user types. For example, the partial map shown in Figure 4 shows two routes, one for each type of user. Whilst the learnt route for the familiar user followed the trajectory 307, 306, 315, 437, 356, 363, 364, the learnt route for the unfamiliar user followed the trajectory 307, 314, 192, 1000, 239, 227, 363, 364. The former route included 28 nodes and is shorter than the latter with 34 nodes, but the latter was found easier to navigate, avoiding user confusions in the stochastic environment.

Finally, additional stochastic behaviour can be incorporated in the spatial environment in order to support other forms of adaptive behaviour. For example, the learning agent can ask the user for about intermediate locations, where

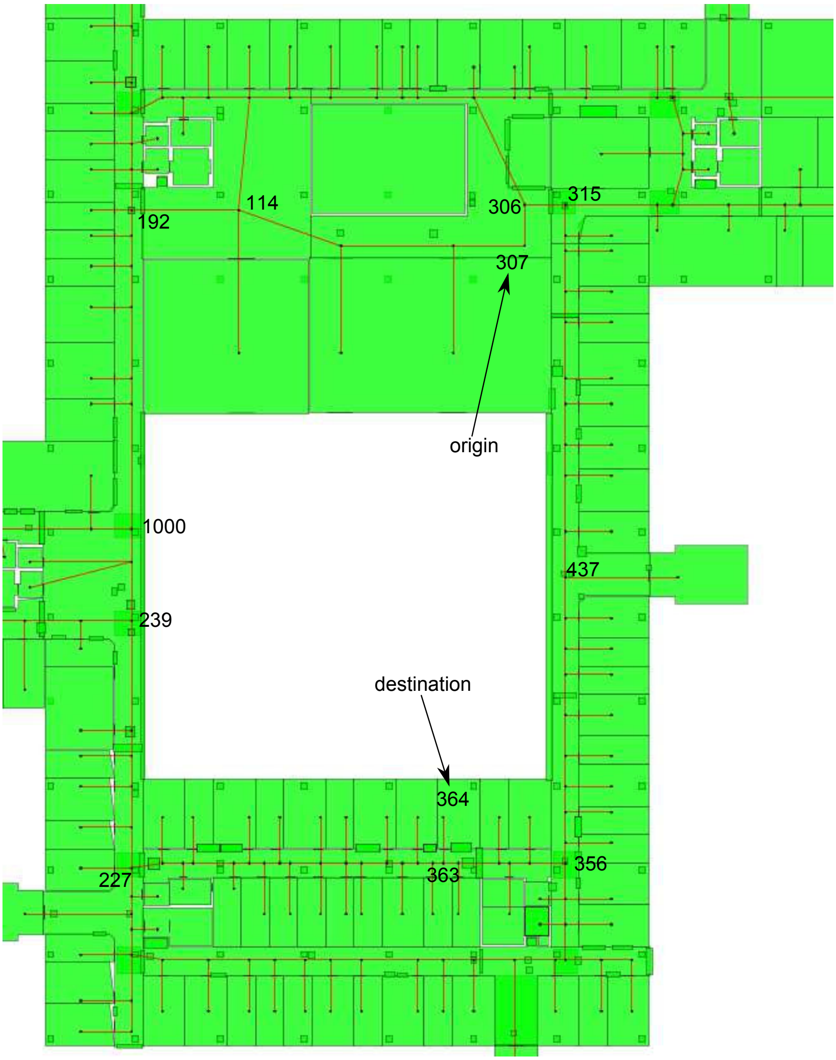


Fig. 4. Partial map illustrating an example of adaptation to the user type: familiar and unfamiliar. Whilst the learnt route for the familiar user followed the trajectory 307, 306, 315, 437, 356, 363, 364, the learnt route for the unfamiliar user followed the trajectory 307, 314, 192, 1000, 239, 227, 363, 364.

each location can incorporate a probability distribution of being known. This is reasonable if we assume that users have partial knowledge of the spatial environment. Intermediate locations may then serve as alternative starting points for providing route directions (e.g., "if you know the main staircase, go there first"). Such probabilistic knowledge can be derived from prior knowledge or learnt from data. The proposed reinforcement learning approach can be used to derive a unified set of optimized behaviours.

6 Conclusions and Future Work

We have presented an approach for fast reinforcement learning of route instruction strategies. This approach decomposes a problem into subproblems in order to learn low-level behaviour in advance (i.e., before user-machine interaction), and at run-time it learns high-level behaviour. We evaluated our approach in a simulated environment for indoor wayfinding, and compared it against flat reinforcement learning and hierarchical fully-learnt behaviour. We found that whilst the latter two require four orders of magnitude (10^4 episodes) to learn the behaviour of an average-sized route, our proposed approach required only three orders of magnitude (10^3 episodes). This result suggests that our method is suitable for its application in online user-machine interactions. Furthermore, the learnt policies induced different behaviour according to different given probabilistic state transition function that modelled user confusions in wayfinding. Thus, the proposed system is able to produce user-adaptive route instructions.

Suggested future work is as follows:

- To evaluate our proposed approach in an environment with real users. This is necessary in order to assess the agents performance in a quantitative and qualitative way. For such a purpose, our proposed approach can be integrated into a (spoken) dialogue system with natural language input and output, as described in [26].
- To investigate how to guide (confused) users in a dialogic interaction, where adaptive guidance would be essential for successful wayfinding. Such behaviour could address the issue of learning to negotiate route instructions. This makes reinforcement learning attractive to wayfinding systems because it optimizes decision-making given the history of the interaction represented in the environment state.
- To test different state representations and reward functions, for instance, the performance function induced from wayfinding dialogues described in [27].
- In a next step we plan to employ techniques of spatial abstraction to include structural knowledge into the state representation [28]. This would allow for another speed-up in learning and for reusing gained knowledge in previously unexplored situations [29]. This is especially useful if the optimization of a hierarchical agent is intractable (i.e., the state-action space becomes too large and is indecomposable).

Acknowledgements

Part of this work was supported by the Transregional Collaborative Research Center SFB/TR 8 "Spatial Cognition." Funding by the German Research Foundation (DFG) is gratefully acknowledged.

References

[1] Lovelace, K.L., Hegarty, M., Montello, D.R.: Elements of good route directions in familiar and unfamiliar environments. In: Freksa, C., Mark, D.M. (eds.) COSIT 1999. LNCS, vol. 1661, pp. 65–82. Springer, Heidelberg (1999)

[2] Sutton, R., Barto, A.: Reinforcement Learing: An Introduction. MIT Press, Cambridge (1998)

[3] Denis, M.: The description of routes: A cognitive approach to the production of spatial discourse. Cahiers Psychologie Cognitive 16(4), 409–458 (1997)

[4] Sorrows, M.E., Hirtle, S.C.: The nature of landmarks for real and electronic spaces. In: Freksa, C., Mark, D.M. (eds.) COSIT 1999. LNCS, vol. 1661, pp. 37–50. Springer, Heidelberg (1999)

[5] Daniel, M.P., Denis, M.: The production of route directions: investigating conditions that favour conciseness in spatial discourse. Applied Cognitive Psychology 18(1), 57–75 (2004)

[6] Klippel, A., Hansen, S., Richter, K.F., Winter, S.: Urban granularities - a data structure for cognitively ergonomic route directions. GeoInformatica 13(2), 223–247 (2009)

[7] May, A.J., Ross, T., Bayer, S.H., Burnett, G.: Using landmarks to enhance navigation systems: Driver requirements and industrial constraints. In: Proceedings of the 8th World Congress on Intelligent Transport Systems, Sydney, Australia (2001)

[8] Ross, T., May, A., Thompson, S.: The use of landmarks in pedestrian navigation instructions and the effects of context. In: Brewster, S., Dunlop, M. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 300–304. Springer, Heidelberg (2004)

[9] Klippel, A., Tenbrink, T., Montello, D.R.: The role of structure and function in the conceptualization of directions. In: van der Zee, E., Vulchanova, M. (eds.) Motion Encoding in Language and Space. Oxford University Press, Oxford (to appear)

[10] Tenbrink, T., Winter, S.: Variable granularity in route directions. Spatial Cognition and Computation: An Interdisciplinary Journal 9(1), 64–93 (2009)

[11] Duckham, M., Kulik, L.: "Simplest" paths: Automated route selection for navigation. In: Kuhn, W., Worboys, M., Timpf, S. (eds.) COSIT 2003. LNCS, vol. 2825, pp. 169–185. Springer, Heidelberg (2003)

[12] Haque, S., Kulik, L., Klippel, A.: Algorithms for reliable navigation and wayfinding. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, D.R. (eds.) Spatial Cognition 2007. LNCS (LNAI), vol. 4387, pp. 308–326. Springer, Heidelberg (2007)

[13] Richter, K.F., Duckham, M.: Simplest instructions: Finding easy-to-describe routes for navigation. In: Cova, T.J., Miller, H.J., Beard, K., Frank, A.U., Goodchild, M.F. (eds.) GIScience 2008. LNCS, vol. 5266, pp. 274–289. Springer, Heidelberg (2008)

[14] Dale, R., Geldof, S., Prost, J.P.: Using natural language generation in automatic route description. Journal of Research and Practice in Information Technology 37(1), 89–105 (2005)

[15] Richter, K.F.: Context-Specific Route Directions - Generation of Cognitively Motivated Wayfinding Instructions. DisKi 314 / SFB/TR 8 Monographs, vol. 3. IOS Press, Amsterdam (2008)
[16] Klippel, A., Richter, K.F., Hansen, S.: Structural salience as a landmark. In: Workshop Mobile Maps 2005, Salzburg, Austria (2005)
[17] Marciniak, T., Strube, M.: Classification-based generation using TAG. In: Belz, A., Evans, R., Piwek, P. (eds.) INLG 2004. LNCS (LNAI), vol. 3123, pp. 100–109. Springer, Heidelberg (2004)
[18] Marciniak, T., Strube, M.: Modeling and annotating the semantics of route directions. In: Proceedings of the Sixth International Workshop on Computational Semantics (IWCS-6), Tilburg, The Netherlands, pp. 151–162 (2005)
[19] Cleary, J., Trigg, L.: An instance-based learner using an entropic distance measure. In: Proceedings of the 12th International Conference on Machine Learning, Tahoe City, Ca, pp. 108–114 (1995)
[20] Kaelbling, L.P., Littmann, M.L., Moore, A.W.: Reinforcement learning: A survey. Journal of Artificial Intelligence Research 4, 237–285 (1996)
[21] Cuayáhuitl, H.: Hierarchical Reinforcement Learning for Spoken Dialogue Systems. PhD thesis, School of Informatics, University of Edinburgh (January 2009)
[22] Cuayáhuitl, H., Renals, S., Lemon, O., Shimodaira, H.: Evaluation of a hierarchical reinforcement learning spoken dialogue system. Computer Speech and Language 24(2), 395–429 (2010)
[23] Dietterich, T.: Hierarchical reinforcement learning with the MAXQ value function decomposition. Journal of Artificial Intelligence Research 13(1), 227–303 (2000)
[24] Dietterich, T.: An overview of MAXQ hierarchical reinforcement learning. In: Choueiry, B.Y., Walsh, T. (eds.) SARA 2000. LNCS (LNAI), vol. 1864, pp. 26–44. Springer, Heidelberg (2000)
[25] Raubal, M., Winter, S.: Enriching wayfinding instructions with local landmarks. In: Egenhofer, M., Mark, D. (eds.) GIScience 2002. LNCS, vol. 2478, pp. 243–259. Springer, Heidelberg (2002)
[26] Cuayáhuitl, H., Dethlefs, N., Richter, K.F., Tenbrink, T., Bateman, J.: A dialogue system for indoor wayfinding using text-based natural language. In: Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), Iasi, Romania (2010)
[27] Dethlefs, N., Cuayáhuitl, H., Richter, K.F., Andonova, E., Bateman, J.: Evaluating task success in a dialogue system for indoor navigation. In: Proc. of the 14th Workshop on the Semantics and Pragmatics of Dialogue, SemDial (2010)
[28] Frommberger, L., Wolter, D.: Spatial abstraction: Aspectualization, coarsening, and conceptual classification. In: Freksa, C., Newcombe, N.S., Gärdenfors, P., Wölfl, S. (eds.) Spatial Cognition VI. LNCS (LNAI), vol. 5248, pp. 311–327. Springer, Heidelberg (2008)
[29] Frommberger, L.: Situation dependent spatial abstraction in reinforcement learning. In: ICML/UAI/COLT Workshop on Abstraction in Reinforcement Learning, Montreal, Canada (2009)

Can Mirror-Reading Reverse the Flow of Time?

Daniel Casasanto[1,2] and Roberto Bottini[1,3]

[1] Max Planck Institute for Psycholinguistics, Neurobiology of Language Group, Nijmegen, NL
[2] Donders Center for Brain, Cognition, and Behavior, Radboud University, Nijmegen, NL
[3] University of Bergamo, Department of Human Sciences, Bergamo, IT
daniel.casasanto@mpi.nl,
roberto.bottini@unibg.it

Abstract. Across cultures, people conceptualize time as if it flows along a horizontal timeline, but the direction of this implicit timeline is culture-specific: in cultures with left-to-right orthography (e.g., English-speaking cultures) time appears to flow rightward, but in cultures with right-to-left orthography (e.g., Arabic-speaking cultures) time flows leftward. Can orthography influence implicit time representations independent of other cultural and linguistic factors? Native Dutch speakers performed a space-time congruity task with the instructions and stimuli written in either standard Dutch or mirror-reversed Dutch. Participants in the Standard Dutch condition were fastest to judge past-oriented phrases by pressing the left button and future-oriented phrases by pressing the right button. Participants in the Mirror-Reversed Dutch condition showed the opposite pattern of reaction times, consistent with results found previously in native Arabic and Hebrew speakers. These results demonstrate a causal role for writing direction in shaping implicit mental representations of time.

Keywords: Culture, Metaphor, Orthography, Space, Time.

1 Introduction

Space and time are intertwined in the human mind, as they are in the physical world. The theory that people use spatial representations to think about time, first inspired by patterns in metaphorical language [1, 2], is now supported by numerous behavioral and neuroscientific experiments [e.g., 3, 4, 5, 6, 7].

Yet, the way people use space to *talk* about time is not necessarily the same way they use space to *think* about it. In English and many other languages, metaphors suggest that time flows along the sagittal (front-back) axis: deadlines lie *ahead of us* or *behind us*; we can *look forward* to our golden years or *look back* on our salad days. Other languages also make use of the vertical axis to talk about time. In Mandarin Chinese, 'the up month' means a month earlier and 'the down month' a month later [8]. Yet, no known spoken language uses the lateral (left-right) axis to talk about time conventionally, and invented left-right metaphors for time sound nonsensical: Monday comes *before* Tuesday, not *to the left of Tuesday* [9].

Despite the total absence of left-right metaphors in spoken language, there is strong evidence that people implicitly associate time with left-right space. Furthermore, the

C. Hölscher et al. (Eds.): Spatial Cognition VII, LNAI 6222, pp. 335–345, 2010.

direction in which time flows along people's imaginary timeline varies systematically across cultures. In one study, Tversky, Kugelmass, & Winter [10] asked children and adults to place stickers on a page to indicate where *breakfast* and *dinner* should appear relative to the *lunch* sticker, in the middle of the page. Whereas English speakers placed breakfast on the left and dinner on the right of lunch, Arabic speakers preferred the opposite arrangement. Fuhrman and Boroditsky [11] showed a similar pattern in a reaction time (RT) task. English- and Hebrew-speaking participants judged whether the second of two pictures showed an earlier or later stage of an unfolding event. English speakers' judgments were fastest when *earlier* was mapped to the left button and *later* to the right, but Hebrew speakers showed the opposite pattern. Ouellet, et al. [12] asked Spanish and Hebrew speakers to judge auditorily presented words referring to the past or future with either their left or right hand, and found a similar reversal of the lateral space-time mapping across groups.

These experimental data reflect patterns that can be found in spontaneous behavior, as well. When English speakers produce co-speech gestures they tend to use the lateral axis for time, much more often than the sagittal axis [13, see also 14; 9; 15]. Earlier times are on the left and later times on the right of body-centered space. Preliminary data from our lab suggests that Spanish speakers' gestures follow a similar pattern, but Arabic speakers' spontaneous gestures show the reverse mapping (Romàn, Casasanto, Jasmin, & Santiago, in prep).

Across cultures, the direction in which time flows along the mental timeline varies predictably with the orthography of the dominant language: time flows rightward in cultures whose literate members use a left-to-right orthography and leftward in cultures that use a right-to-left orthography. Yet, despite this clear correlation, it is not known to what extent reading and writing direction is a *cause* or an *effect* of cross-cultural variation in implicit space-time mappings.

In principle, a culture's writing system could emerge with one directionality or another as a consequence of culture-specific conceptions of time -- not the other way around. This seems especially plausible for cultures where literacy (or mass-literacy) is a recent development. Alternatively, directionality in *both* orthography and in thought could arise due to cultural bootstrapping from material artifacts like calendars (whether the calendar is a grid on a piece of paper, knots on a string, notches on a branch, etc.) or other devices for keeping track of time (e.g., a solar clock) or number (e.g., a horizontal abacus) [16]. Cultural practices tend to covary: groups who write from left to right often spatialize time on calendars and numbers on graphs from left to right, as well. Based on correlational data, it is not possible to determine whether experience reading or writing plays any causal role in fixing the direction of implicit space-time mappings.

Here we performed an experimental intervention to determine whether experience with reading a left-to-right or right-to-left orthography is sufficient to determine the direction of people's implicit associations from space to time. Native Dutch speakers were assigned to perform one of two space-time congruity tasks. In one task (Experiment 1), participants saw past-oriented phrases (e.g. *a year earlier*) and future-oriented phrases (e.g. *a decade later*) appear on the screen one at a time, in standard Dutch orthography. As soon as each phrase appeared, they pressed a button (located on the left or right of a keyboard) to indicate the temporal reference of the phrase (past or future). Each participant performed two blocks: in one block the left-right key mapping required responses that were congruent with a left-to-right flow of time, and in the other responses were congruent with a right-to-left mapping. The order of blocks was counterbalanced across participants. We predicted that, on average, participants would show an RT advantage for responses consistent with standard Dutch orthography (left-to-right).

The other task (Experiment 2) was identical to the first, with one exception: all instructions and stimuli were presented in mirror-reversed text. Reading requires scanning the page in a particular direction: normally for Dutch speakers reading each line of a text requires moving the eyes gradually from the left to the right side of the page or the computer screen. As such, moving rightward in space is tightly coupled with 'moving' later in time. We reasoned that if the habit of reading from left-to-right contributes to an implicit left-to-right mapping of time in readers' minds, then practice reading in the opposite direction should weaken and eventually reverse this mapping.

2 Space-Time Congruity with Standard Orthography

In Experiment 1, all instructions and stimuli were presented in standard Dutch orthography. We conducted Experiment 1 to validate the use of this space-time congruity paradigm in native Dutch speakers, and to provide a comparison group for the mirror-reading group.

2.1 Methods

Participants. Native Dutch speakers (N=32) performed Experiment 1 in exchange for payment.

Stimuli. Temporal phrases were constructed in Dutch, each with 3 words. The first word was an indefinite article, the second word a temporal interval, (tr., *second*, *moment*, *minute*, *hour*, *day*, *week*, *month*, *season*, *year*, *decade*, *century*, *millennium*), and the third word a temporal modifier (tr., *before*, *after*, *earlier*, *later*). The twelve temporal intervals were fully crossed with the four temporal modifiers to produce 48 temporal phrases (e.g., *a day before*; *a century after*; *a year earlier*; *a week later*). Half of the phrases referred to an earlier (past-oriented) interval of time (i.e., if the modifier was *earlier* or *before*), and the other half referred to a later (future-oriented) interval (i.e., if the modifier was *later* or *after*). Two of the modifiers were spatial terms used metaphorically (*before*, *after*), and the other two were purely temporal terms with similar meanings (*earlier*, *later*). Phrases were presented in the center of a Macintosh laptop screen (resolution=1024x768), in black 48-point Arial font, on a white background.

Apparatus. Participants were seated at a desk. Two A4 Xerox paper boxes were stacked on the desk, and a laptop computer was secured on top of them, to raise the screen to approximately the participants' eye-level. A standard USB keyboard was mounted horizontally on the side of the upper box, with the keys facing the participant, at about shoulder level. The keyboard was covered with a sheet of black plastic with holes that exposed only the three keys needed for responses: the "A" key on the left, the "apostrophe" key on the right, and the "H" hey in the middle. The middle key was aligned with the center of the laptop screen, and the left and right keys were equidistant from it. The left key was covered with a blue sticker and the right key a red sticker, or vice versa, with the key colors counterbalanced across subjects.

Procedure. The experiment consisted of two blocks. In each block, each of the 48 temporal phrases was presented once, for a total of 96 trials. Written instructions appeared on the screen before each block. In one of the blocks, participants were instructed that as soon as each phrase appeared, they should press the blue button if the phrase referred to an interval of time in the past (e.g., a week *earlier*) and the red button if it referred to an interval of time in the future (e.g., a week *later*). In the other block, the mapping between the red/blue keys and pastward/futureward phrases was reversed. To ensure that participants remembered the correct color-time mapping, after reading the instructions they were required to rehearse the correct color-time mapping aloud 5 times, before each block (e.g., "past=blue, past=blue, past=blue, etc.; future=red, future=red, future=red, etc.)

At the beginning of each trial the word 'ready' appeared in the center of the screen and remained there until the participant pressed the middle white button. 'Ready' was then replaced by a fixation cross. Participants were instructed to hold down the white button for as long as the fixation was shown. Its duration was varied randomly from 300-450 ms, in 50 ms increments, to make its duration unpredictable and discourage anticipatory movements. The fixation was then replaced by one of the 48 temporal phrases. Participants were instructed to press the colored button corresponding to the temporal reference of the phrase as quickly and accurately as possible. The phrase remained on the screen until the participant responded, at which time it was replaced by the 'ready' message to begin the next trial.

Participants pressed buttons with the index finger of the dominant (writing) hand. To ensure they would used the same hand for both rightward and leftward responses, participants were required to sit on their non-dominant hand.

The spatial direction of responses was never mentioned, but one colored button was on the right and the other on the left of the middle white button. Therefore, in one block pressing the correctly colored button called for a movement that was congruent with the space-time mapping encoded in standard Dutch orthography (e.g., pressing the blue button for a pastward phrase when the blue button was on the left); in the other block pressing the correctly colored button called for an incongruent movement (e.g., pressing the blue button for a futureward phrase when the blue button was on the left). The order of congruent-movement and incongruent-movement blocks was counterbalanced across participants. The space-time congruity effect was computed for each subject by comparing response times during Congruent and Incongruent responses (between-blocks, within-items). Testing lasted about 10 minutes.

2.2 Results and Discussion

Participants pressed the correct button on 96% of trials. Only accurate responses were analyzed. This resulted in the removal of 4% of the data. Responses greater than 5000 ms were also excluded, which resulted in the removal of 0.2% of the accurate trials.

A 2 X 2 ANOVA was conducted with Congruity of Movement Direction (Congruent with Time flowing leftward, Congruent with time flowing rightward) and Block (Block 1, Block 2) as within-subject and within-item factors. There was a highly significant main effect of Congruity ($F_1(1,15)=30.56$, $p=.0001$; $F_2(1,47)=119.38$, $p=.0001$). There was no main effect of Block (F_1(1,15)=0.75, *ns*; F_2(1,47)=2.99, *ns*). The Congruity X Block interaction was significant by items but not by subjects (F_1(1,15)=1.41, *ns*; F_2(1,47)=6.27, *p*=.02).

Congruity of Movement was then compared within each block (Block 1: F_1(1,30)=9.62, *p*=.004; F_2(1,47)=116.31, *p*=.0001; Block 2: F_1(1,30)=3.64, *p*=.07; F_2(1,47)=32.55, *p*=.0001). Mean RTs are shown in figure 1.

Overall, there was a strong effect of Congruity. Participants responded faster when the mapping between the color of the buttons and the temporal reference of the phrases required leftward movements for past-oriented phrases and rightward movements future-oriented phrases. This space-time congruity effect is similar to effects found previously in English and Spanish speakers (e.g., Torralbo, et al., 2006; Weger & Pratt, 2008). We are not aware of previous studies showing this effect in Dutch speakers, but given the correlation between writing direction and the direction of the space-time mappings across cultures, we had no reason to expect that Dutch speakers should perform differently from speakers of other languages that use a Roman alphabet.

For our present purposes, it is important that this paradigm produced a congruity effect in the same direction for both blocks. Having shown that this task provides clear evidence for the implicit space-time mapping typically found in left-to-right reading cultures, we can proceed to test effects of exposure to an orthography in which 'progress' along a spatio-temporal continuum proceeds in the opposite direction.

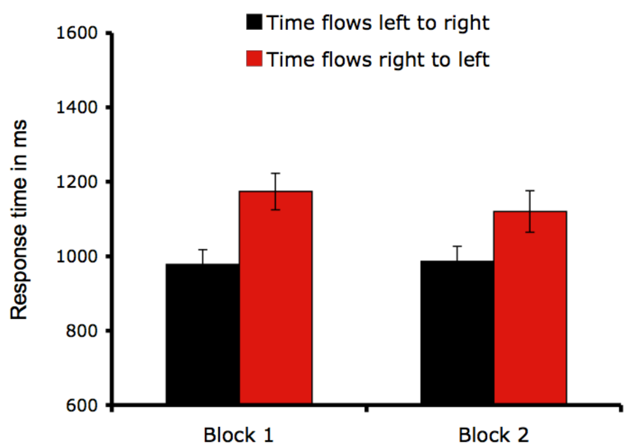


Fig. 1. Results of Experiment 1. When reading standard orthography, participants were faster to judge the temporal reference of phrases when the left-right button press responses were consistent with the rightward flow of time along an implicit mental timeline (left=earlier, right=later). Error bars indicate s.e.m.

3 Space-Time Congruity with Mirror-Reversed Orthography

To test for a causal role of orthography in the mental representation of temporal order, we replicated Experiment 1 in a new group of participants using stimuli and instructions presented in mirror-reversed font.

3.1 Methods

Participants. A new sample of native Dutch speakers (N=32) performed Experiment 2 in exchange for payment.

Materials and Procedure. Materials and procedures were identical to Experiment 1, with one exception. All instructions and stimuli were presented mirror-reversed. Testing lasted about 15 minutes.

3.2 Results and Discussion

Participants pressed the correct button on 97% of trials. Only accurate responses were analyzed. This resulted in the removal of 3% of the data. Responses greater than 5000 ms were also excluded, which resulted in the removal of 4% of the accurate trials trials.

A 2 X 2 ANOVA was conducted with Congruity of Movement Direction (Congruent with Time flowing leftward, Congruent with time flowing rightward) and Block (Block 1, Block 2) as within-subject and within-item factors. There was no main effect of Congruity ($F_1(1,15)=.79$, *ns*; $F_2(1,47)= 2.29$, *ns*). There was a highly significant effect of Block ($F_1(1,15)= 66.37$, $p=.0001$; $F_2(1,47)= 321.81$, $p=.0001$), and crucially, a highly significant Congruity X Block interaction ($F_1(1,15)= 31.89$, $p=.0001$; $F_2(1,47)= 206.56$, $p=.0001$).

Congruity of Movement was then compared within each block (Block 1: $F_1(1,30)=5.00$, $p=.03$; $F_2(1,47)=98.36$, $p=.0001$; Block 2: $F_1(1,30)= 3.02$, $p.=.09$; $F_2(1,47)= 125.21$, $p=.0001$). Mean RTs are shown in figure 2.

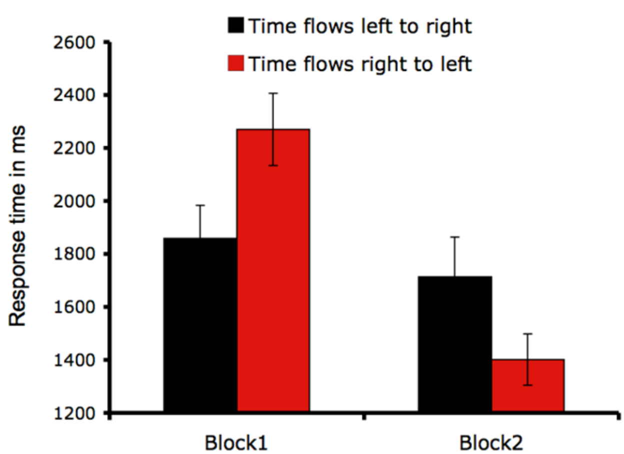


Fig. 2. Results of Experiment 2. In the first block, the space-time congruity effect was similar to that observed in Experiment 1. But by the second block of reading mirror-reversed orthography, this congruity effect reversed. Participants were faster to judge the temporal reference of phrases when the left-right button press responses were consistent with a leftward flow of time along an implicit mental timeline (left=later, right=earlier). Error bars indicate s.e.m.

Finally, we compared the congruity effects found in Experiment 1 and Experiment 2 using a 2 X 2 X 2 ANOVA with Congruity and Block as within-subject/within-item factors and Orthography (Standard orthography, Mirror-reversed orthography) as a within-subject/within-item factor. Consistent with the prediction that orthography can influence mental representations of time, we find a highly significant 3-way interaction ($F_1(1,30)= 22.71$, $p=.0001$; $F_2(1,94)=125.38$, $p=.0001$). By subtracting the RTs during trials where movements were congruent with the leftward flow of time from RTs during trials where movements were congruent with the rightward flow of time (RT_rightward - RT_leftward), this 3-way interaction can be simplified, and conceptualized as a 2-way interaction of Block X Orthography (see figure 3).

As is evident from figure 3, the absolute values (ABS) of the congruity effects in both blocks of Experiment 2 are greater than the ABS of the effects in Experiment 1. This was not expected, and although it is not relevant to our experimental hypothesis, it bears further investigation. On one possible explanation, congruity effects may result from a failure of cognitive control; that is, they result from participants' inability to ignore the irrelevant spatial dimension of their responses when judging the temporal reference of the stimuli. The cross-dimensional effect of space on time judgments may have been greater in Experiment 2 because cognitive control resources were taxed by reading backwards, contrary to habit.

Although the dominant space-time mapping in Dutch culture continued to influence RTs during the first block of Experiment 2, by the second block exposure to mirror-reversed writing was sufficient to reverse the congruity effect. Since this is the first experiment to test for a causal influence of writing direction on time representation, we did not have any *a priori* prediction about how much experience with

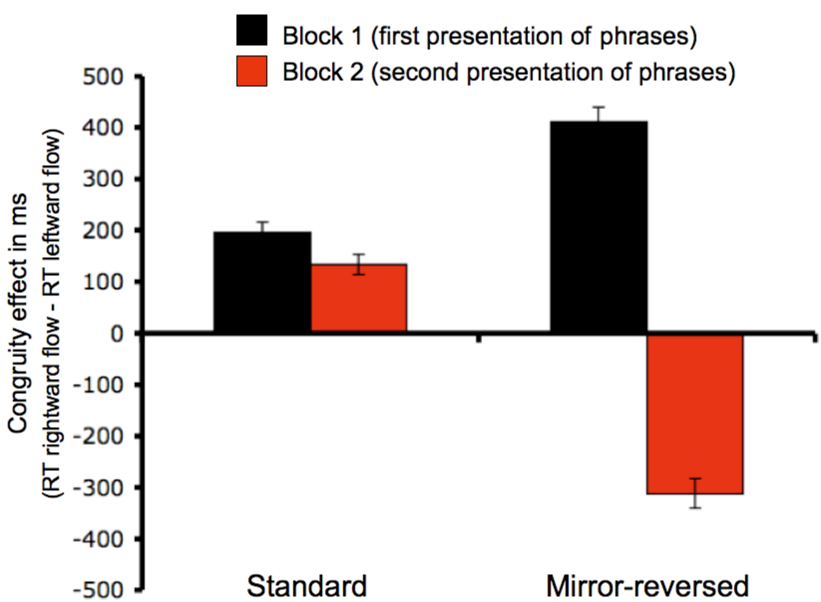


Fig. 3. Congruity effects (RT_rightward - RT_leftward) across blocks for Experiment 1 (left) and Experiment 2 (right). Error bars indicate s.e.m.

reversed orthography would be needed to produce a significant change in the congruity effect, nor could we predict whether the congruity effect would be reversed or merely diminished. To support our hypothesis, it would have been sufficient to show a reduction in the left-to-right congruity effect from Block 1 to Block 2 that was greater in Experiment 2 than in Experiment 1. However, the fact that the congruity effect completely reversed here provides a particularly clear demonstration that even brief experience with one orthography or another can influence people's implicit spatial representations of time. [For compatible evidence of the flexibility of space-time metaphors in language and thought, see 4, 8, 17, 1; 18, 10].

4 General Discussion

It is now well established that people activate implicit associations between space and time when processing temporal language, and that the specifics of these associations vary systematically across cultures [11, 12, 10]. Since time is not associated with left-right space in any known linguistic metaphors, it is unlikely that these culture-specific mappings are learned through experience with spoken language.[1] Here we tested whether orthography can play a causal role in fixing the direction in which time flows along the imaginary mental timeline. Experiment 1 showed that, when exposed to temporal phrases presented in standard left-to right orthography, Dutch speakers implicitly associated earlier time intervals with leftward movements and later time intervals with rightward movements, consistent with previous findings in members of other cultures that use the Roman writing system.

However, when exposed to several minutes of mirror-reversed writing, Dutch participants began to show space-time congruity effects that revealed a reversal of their normally dominant implicit space-time mapping. By the second time they were judging each of the 48 temporal phrases (Block 2 of Experiment 2), participants were faster to make responses when key presses associated earlier events with *rightward* movements and later events with *leftward* movements -- a pattern observed previously in speakers of Hebrew, which is written from right to left. It appears that experience reading a right-to-left orthography (which requires the reader to 'progress' leftward across the screen with his/her eyes) is sufficient to reverse the flow of time in the reader's mind, at least transiently.

Although this rapid retraining of a space-time association stored in long-term memory may seem surprising, it is not unprecedented. In one study, Boroditsky [8] found that horizontal spatial primes facilitated English speakers' judgments of temporal sentences (e.g., *April comes earlier than May*) more than vertical primes did, but found the opposite pattern in Mandarin speakers, consistent with the difference between these languages in the prevalence of horizontal and vertical metaphors for time. To test whether linguistic experience could affect these mappings, she trained a new group of English speakers to use Mandarin-like vertical spatial metaphors for time. After brief training, English speakers showed a pattern of priming similar to native Mandarin speakers.

---

[1] Although spoken languages do not use the lateral axis for time, some signed languages do [23].

In a test of a different set of space-time metaphors Casasanto [17] and colleagues showed that when English and Greek speakers perform non-linguistic duration reproduction tasks, they show language-specific patterns of cross-dimensional interference from space. Whereas English speakers have a harder time screening out interference from (1-dimensional) spatial distance, Greek speakers have more difficulty screening out interference of (3-dimensional) volume. This pattern was predicted based on the relative prevalence and productivity of distance and volume metaphors for duration across languages (e.g., a *long* time (like a *long* rope); *a large amount of* time (like *a large amount* of water)). To find out whether using volume metaphors could cause the volume-interference found in Greeks, US English speakers were trained to use Greek-like volume metaphors for time. Results showed that after one brief (but concentrated) training session, English participants showed a pattern of cross-dimensional interference from volume in a low-level psychophysical task that was statistically indistinguishable from the pattern seen in native Greek speakers.

Time is not the only domain that appears to be mentally represented, in part, through spatial metaphors (which may or may not correspond to linguistic metaphors). Emotional valence is also spatialized on a left-right axis: whereas right-handers tend to associate the right hand and the right side of space with positive things and the left with bad, left-handers show the opposite set of implicit associations [19]. It was proposed that this mapping arises due to asymmetries in motor fluency: people like things on their dominant side better because they can interact with things on that side more easily. To test this proposal, Casasanto [20] asked right-handers to perform a 2-part training task. In the first part, they arranged dominoes according to a symmetrical pattern on a tabletop, standing them on end, moving both hands in synchrony. The challenge was that they were randomly assigned to wear a bulky ski glove one hand or the other while performing the task, which either enhanced their natural right-handedness or made them temporarily more skillful with their left hand.

After 12 minutes of this asymmetric motor experience, participants were taken to a different room by a different experimenter for some ostensibly unrelated questionnaire studies, one of which tested implicit associations between space and valence. This questionnaire was shown previously to produce distinctive patterns of judgments in right- and left-handers [19]. Participants whose training experience preserved their natural dominance showed the typical right-handers' pattern. But participants who had worn the skiglove on their right hand during training, becoming transiently left-handed, produced a pattern of responses that was indistinguishable from natural lefties'.

We are aware of one training study that manipulated writing direction in order to test a role for orthography in the spatial representation of gender and agency. Several studies suggest that males (seen as more agentive) tend to be represented to the left of females in the minds of people who speak left-to-right languages like English, but not for speakers of right-to-left languages like Arabic [21]. Yet, Suitner [21] showed that this spatial bias can be nullified in speakers of Italian who are trained to perform a leftward writing exercise, reversing not only their habitual writing direction but also their habitual associations of gender, agency, and space.

How enduring are these training effects? Presumably, without further reinforcement of the new habits, participants who show rapid training effects will also revert to their long-term habits rapidly. Exactly how soon remains a question for further research. Depending on the goal of the training manipulation, the durability of the behavioral change may matter more or less. In the present study the goal was to test the sufficiency of a proposed cause of cross-cultural differences. The total reversal of the congruity effect as a function of reading experience demonstrates that orthography can, indeed, influence the implicit spatial representation of time. This simple demonstration that orthography can play a causal role in directing the flow of time along the mental timeline would serve its theoretical goal even if the effect were quickly reversed when participants resumed normal reading habits.

How best to characterize the learning mechanisms that afford this representational plasticity remains another open question. It may be fruitful to consider the changes participants undergo in Experiment 2 in terms of a hierarchical Bayesian model [22]. To sketch this suggestion briefly, people's associations between space and time could be characterized as intuitive hypotheses. Based on ordinary reading experience, Dutch speakers form the hypothesis that by default events unfold from left to right. Yet after training, they appear to entertain the hypothesis that events unfold from right to left.

To explain how participants can switch from one hypothesis to a contradictory hypothesis (and presumably switch back) so quickly, it may help to posit that they also entertain a more enduring overhypothesis, of which both the 'Dutch-like' and 'Arabic-like' space-time associations are specific instances. The overhypothesis could be that time is associated with motion along a linear path. Such a belief would be well supported by observable correlations in the physical world: spatial succession is a reliable index of temporal succession.

Consistent with this proposal, we suggest that if orthography is responsible for determining the direction in which time flows along people's left-right mental timelines, this directional mapping likely builds upon a prior less-specific space-time association, which arises (either in developmental or evolutionary time) from space-time correlations that have no particular directionality: on any trajectory, it is the case that as a moving object travels farther, more time passes. The hierarchical model can potentially help to explain how 'mental metaphors' linking space-time can be universal at one level of level of description but culture-specific at another.

Acknowledgments. This research was supported in part by the Max Planck Gesellschaft and by an NRSA fellowship (#F32MH072502) and by a grant from the Spanish Ministry of Education and Science (#SEJ2006-04732/PSIC, DGI) to DC.

References

1. Clark, H.H.: Space, time, semantics and the child. In: Moore, T.E. (ed.) Cognitive Development and the Acquisition of Language, pp. 27–63. Academic Press, New York (1973)
2. Lakoff, G., Johnson, M.: Metaphors We Live by. The Chicago University Press, Chicago (1980)
3. Basso, G., Nichelli, P., Frassinetti, F., di Pellegrino, G.: Time perception in a neglected space. Neuroreport 7, 2111–2114 (1996)
4. Boroditsky, L.: Metaphoric structuring: understanding time through spatial metaphors. Cognition 75(1), 1–28 (2000)
5. Casasanto, D., Boroditsky, L.: Time in the Mind: Using space to think about time. Cognition 106, 579–593 (2008)

6. Torralbo, A., Santiago, J., Lupiáñez, J.: Flexible conceptual projection of time onto spatial frames of reference. Cog. Sci. 30, 749–757 (2006)
7. Weger, U.W., Pratt, J.: Time flies like an arrow: Space-time compatibility effects suggest the use of a mental time-line. Psychonomic Bulletin & Review 15, 426–430 (2008)
8. Boroditsky, L.: Does language shape thought? English and Mandarin speakers' conceptions of time. Cog. Psych. 43(1), 1–22 (2001)
9. Cienki, A.: Metaphoric gestures and some of their relations to verbal metaphoric expressions. In: Koening, J.-P. (ed.) Discourse and cognition: Bridging the gap, pp. 189–204. CSLI Publications, Stanford (1998)
10. Tversky, B., Kugelmass, S., Winter, A.: Cross-cultural and developmental trends in graphic productions. Cog. Psych. 23, 515–557 (1991)
11. Fuhrman, O., Boroditsky, L.: Mental time-lines follow writing direction: Comparing English and Hebrew speakers. In: Proceedings of 29th Annual Conference of the Cognitive Science Society, Nashville, TN (2007)
12. Ouellet, M., Santiago, J., Israeli, Z., Gabay, S.: Is the future the right time? Exp. Psych. (in press)
13. Casasanto, D.: When is a linguistic metaphor a conceptual metaphor? In: Evans, V., Pourcel, S. (eds.) New Directions in Cognitive Linguistics, pp. 127–145. John Benjamins, Amsterdam (2009)
14. Boroditsky, L.: Do English and Mandarin Speakers Think Differently About Time? In: Proceedings of the 31th Annual Conference of the Cognitive Science Society, pp. 575–580. Lawrence Erlbaum Associates, Hillsdale (2008)
15. Cooperrider, K., Nunez, R.: Across time, across the body Transversal temporal gestures. Gesture 9, 181–206 (2009)
16. Dehaene, S.: The number sense. Oxford University Press, New York (1999)
17. Casasanto, D.: Who's afraid of the Big Bad Whorf? Cross-linguistic differences in temporal language and thought. Language Learning 58(1), 63–79 (2008)
18. Evans, V.: The structure of time: Language, meaning and temporal cognition. John Benjamins, Amsterdam (2004)
19. Casasanto, D.: Embodiment of Abstract Concepts: Good and bad in right- and left-handers. JEP: General 138(3), 351–367 (2009)
20. Casasanto, D.: Motor experience shapes abstract concepts. In: Proceedings of the 50th Annual Meeting of the Psychonomic Society, Boston, MA (2009)
21. Suitner, C.: Where to place social targets? Stereotyping and Spatial Agency Bias. Doctoral dissertation, Department of Psychology, University of Padova (2009)
22. Kemp, C., Perfors, A., Tenenbaum, J.B.: Learning ovehypotheses with hierarchical Bayesian models. Dev. Science 10(3), 307–321 (2007)
23. Emmorey, K.: Space on hand: The exploitation of signing space to illustrate abstract thought. In: Gattis, M. (ed.) Spatial schemas and abstract thought, pp. 147–174. The MIT Press, Cambridge (2001)

Author Index