

**The Psychology of
Learning and Motivation**

Volume 58





VOLUME FIFTY - EIGHT

THE PSYCHOLOGY OF LEARNING AND MOTIVATION

Series Editor

BRIAN H. ROSS

*Beckman Institute and Department of Psychology
University of Illinois, Urbana, Illinois*



VOLUME FIFTY - EIGHT

THE PSYCHOLOGY OF LEARNING AND MOTIVATION

Edited by

BRIAN H. ROSS

*Beckman Institute and Department of Psychology
University of Illinois, Urbana, Illinois*



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford, OX51GB, UK
32, Jamestown Road, London NW1 7BY, UK

Copyright © 2013 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at <http://elsevier.com/locate/permissions>, and selecting *Obtaining permission to use Elsevier material*

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

ISBN: 978-0-12-407237-4

ISSN: 0079-7421

For information on all Academic Press publications
visit our website at store.elsevier.com

Printed and bound in USA

13 14 15 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

CONTENTS

Contributors

vii

1. Learning Along With Others	1
Robert L. Goldstone, Thomas N. Wisdom, Michael E. Roberts, Seth Frey	
1. Learning within a Community of Learners	2
2. Innovation Propagation in a One-dimensional Problem Space	9
3. Collective Learning in Higher-dimensional Problem Spaces	24
4. Collective Search in a Problem Space with Interactions among Solution Elements	31
5. Limitations, Implications, and Conclusions	37
2. Space, Time, and Story	47
Barbara Tversky, Julie Heiser, Julie Morrison	
1. Introduction: Life, Perception, and Stories	47
2. Perceiving the Events of Life	48
3. Telling the Events of Life: Word, Diagram, Gesture	53
4. Spraction: Space, Action, Abstraction	72
3. The Cognition of Spatial Cognition: Domain-General within Domain-specific	77
Holly A. Taylor, Tad T. Brunyé	
1. Spatial Cognition: Exploring the Domain-general in the Domain-specific	78
2. Spatial Cognition and Attention	79
3. Memory and Spatial Cognition	91
4. Representing Spatial Information: Embodiment	96
5. The Cognition of Spatial Cognition	108
4. Perceptual Learning, Cognition, and Expertise	117
Philip J. Kellman, Christine M. Massey	
1. Introduction	118
2. Perceptual Learning Effects	122
3. Perceptual Learning in Mathematics: An Example	124
4. Perception, Cognition, and Learning	130
5. Perceptual Learning and Instruction	144
6. Conclusion	159

5. Causation, Touch, and the Perception of Force	167
Phillip Wolff, Jason Shepard	
1. Introduction	168
2. Force-based Accounts of Causation	169
3. Grounding Forces in the Sense of Touch	174
4. Evidence for a Force-based View of Causation from Work on the Perception of Causation	178
5. The Sense of Force in Causal Perception and Induction	183
6. The Feeling of Causation in the Absence of Mechanism: A Dual Process Approach	191
7. Potential Problems for a Force-based View of Causation	196
8. Conclusions	198
6. Categorization as Causal Explanation: Discounting and Augmenting in a Bayesian Framework	203
Daniel M. Oppenheimer, Joshua B. Tenenbaum, Tevye R. Krynski	
1. Introductory Comments	204
2. Previous Accounts of Categorization	206
3. The Present Model	212
4. Tests of the Predictions	217
5. Placing the Present Account in the Literature	223
6. Conclusion	228
7. Individual Differences in Intelligence and Working Memory: A Review of Latent Variable Models	233
Andrew R. A. Conway, Kristof Kovacs	
1. Introduction	234
2. Models of Intelligence	240
3. Interpretation of the General Factor	247
4. Models of Working Memory Capacity	252
5. Toward a Unified Model	256
6. Conclusion	262
Appendix	263
<i>Index</i>	<i>271</i>

CONTRIBUTORS

Tad T. Brunyé

Department of Psychology, Tufts University, Medford, MA, USA
U.S. Army Natick Soldier Research, Development, and Engineering Center, Natick, MA, USA

Andrew R.A. Conway

Department of Psychology, Princeton University, Princeton, NJ, USA

Seth Frey

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Robert L. Goldstone

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Julie Heiser

Adobe Systems, San Jose, CA

Phillip J. Kellman

Department of Psychology, University of California, Los Angeles, CA, USA

Kristof Kovacs

Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

Tevye R. Krynski

Department of Brain and Cognitive Sciences, M.I.T., Cambridge, MA, USA

Christine M. Massey

Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA, USA

Julie Morrison

Department of Psychology, Glendale Community College, Glendale AZ

Daniel M. Oppenheimer

Department of Psychology, Princeton University, Princeton, NJ, USA
Anderson School of Management, University of California, Los Angeles, CA, USA

Michael E. Roberts

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Jason Shepard

Department of Psychology, Emory University, Atlanta, GA, USA

Holly A. Taylor

Department of Psychology, Tufts University, Medford, MA, USA

Joshua B. Tenenbaum

Department of Brain and Cognitive Sciences, M.I.T., Cambridge, MA, USA

Barbara Tversky

Teachers College, Columbia University, New York, NY, USA

Department of Psychology, Stanford University, Palo Alto, CA, USA

Thomas N. Wisdom

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Phillip Wolff

Department of Psychology, Emory University, Atlanta, GA, USA



Learning Along With Others

**Robert L. Goldstone¹, Thomas N. Wisdom, Michael E. Roberts,
Seth Frey**

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

¹Corresponding author: E-mail: rgoldsto@indiana.edu

Contents

1. Learning within a Community of Learners	2
1.1. Parallel but Interactive Learning in Groups	4
1.2. Bridging between Individual and Group Levels of Explanation	6
1.3. Exploration and Exploitation	7
2. Innovation Propagation in a One-dimensional Problem Space	9
2.1. An Experimental Examination of Connectedness and Fitness Functions	11
2.2. A Computational Model of Innovation Propagation	18
3. Collective Learning in Higher-dimensional Problem Spaces	24
3.1. The "Draw the Mystery Picture" Task	25
3.2. Major Results and Implications	27
4. Collective Search in a Problem Space with Interactions among Solution Elements	31
4.1. The "Creature League" Task	32
4.2. Major Results and Implications	34
5. Limitations, Implications, and Conclusions	37
5.1. Imitation Heuristics	38
5.1.1. Frequency	38
5.1.2. Upgrade	38
5.1.3. Early Imitation	38
5.1.4. Similarity	39
5.1.5. Momentum	39
5.2. Group-level Phenomena	40
5.2.1. Convergence	40
5.2.2. Inefficient Problem Space Coverage	41
5.2.3. Problem Space/Social Knowledge Match	41
5.2.4. Reciprocal Imitation	41

Abstract

Unlike how most psychology experiments on learning operate, people learning to do a task typically do so in the context of other people learning to do the same task. In these situations, people take advantage of others' solutions, and may modify and extend these solutions, thereby affecting the solutions available to others. We are interested in

the group patterns that emerge when people can see and imitate the solutions, innovations, and choices of their peers over several rounds. In one series of experiments and computer simulations, we find that there is a systematic relation between the difficulty of a problem search space and the optimal social network for transmitting solutions. As the difficulty of finding optimal solutions in a search space increases, communication networks that preserve spatial neighborhoods perform best. Restricting people's access to others' solutions can help the group as a whole find good, hard-to-discover solutions. In other experiments with more complex search spaces, we find evidence for several heuristics governing individuals' decisions to imitate: imitating prevalent options, imitating options that become increasingly prevalent, imitating high-scoring options, imitating during the early stages of a multiround search process, and imitating solutions similar to one's own solution. Individuals who imitate tend to perform well, and more surprisingly, individuals also perform well when they are in groups with other individuals who imitate frequently. Taken together, our experiments on collective social learning reveal laboratory equivalents of prevalent social phenomena such as bandwagons, strategy convergence, inefficiencies in the collective coverage of a problem space, social dilemmas in exploration/exploitation, and reciprocal imitation.



1. LEARNING WITHIN A COMMUNITY OF LEARNERS

From the procedures that many psychology experiments on learning use, one might get the impression that learning is typically a solitary affair. Experiment participants are often given an inductive learning task to perform in the seclusion of their own cubicle, with a minimum of instructions or advice. Participants are isolated from each other for reasons of good experimental control. If participants were able to “look over each others' shoulders” to see how others are solving the task, they might adopt their solutions, and cease to be an independent data point.

In psychology, one researcher's confound is another researcher's object of study. We control variables because we expect them to exert a potentially large, contaminating influence on the topic under study. Psychologists may throw out the first 200 trials of a 2000 trial experiment because they want to observe stable performance, not performance affected by learning. Psychologists counterbalance the ordering of trials because of context, learning, and motivational effects. They run double-blind experiments to control for expectancies, goal-driven perception, and social influence. Of course, all of these experimental artifacts—context, learning, motivation, expectancies, motivated perception, and social influence—are also potent and important psychological phenomena deserving inquiry on their own.

We believe that social learning is another such psychological phenomenon. By purposefully allowing participants to peer over their peers'

shoulders, allowing structured communication between participants trying to solve the same task, we can increase our understanding of how human learning often occurs.

As a species, humans are “obligatorily gregarious,” to use the zoo classification for species in which the individuals do not thrive unless they are living in a group of their own kind (Cacioppo & Patrick, 2008). We are typically surrounded by other people, and our degree of connectivity is rapidly increasing with the growing Internet, the increasing prevalence of mobile networked devices, and decreasing travel costs (Goldstone & Gureckis, 2009). As we try to solve problems in our everyday life, solutions from other people are readily at hand. This can be unfortunate when we are trying to form our own opinion about a movie, see an old television series without having its ending spoiled, or solve a difficult puzzle on our own without giving in to the sirens’ call of online solution sites. More often, though, we solve problems much better because we have access to others’ solutions.

A striking example of this is the speed with which software developers can now create highly sophisticated computer programs. The “open source software” community is committed to making software products, including the source code for the software, available to any interested party without restrictions (Lerner, Tirole, & Pathak, 2006). Due in part to this vibrant community, programmers now have a veritable smorgasbord of packages and libraries at their disposal when they are adding their own contributions to this collective repository. Previous software solutions are tweaked, adapted, and generalized to fit new needs, and developers frequently find it reinforcing, not aversive, when other developers use their solutions. Scientific progress in academic settings typically works in a similar fashion, with scientists benefiting tremendously from being in a community of other scientists who are making their methods, tools, analyses, theories, and experimental results available to others (Simon, 1957). In software development and science, not to mention music, art, sports, medicine, farming, and government, progress is radically expedited by innovators leveraging the work of others, learning from, and extending, previous solutions.

Outside of psychology experiment cubicles, learning typically takes place in a community of learners. Accordingly, we are interested in bringing back to the cubicles some of the essential elements of social learning. As experimental psychologists, we are loath to throw out the experimental control baby with the assumption of isolated learning bathwater. Our *modus operandi* has been to allow participants in laboratory experiments to

view each others' solutions and then to imitate and modify these solutions if they so choose. We do not allow participants to see or have open-ended conversations. Although these higher bandwidth channels of communication have produced important results (Ostrom, Gardner, & Walker, 1994), they are less amenable to the kind of computational modeling we develop in Section 3. Our participants have highly constrained communication possibilities. They can only view each others' solutions and the scores earned by those solutions. However, this minimal information exchange is still sufficient for creating emergent patterns of group convergence and collective coverage of a problem space that are commonly observed in real groups of interacting problem solvers.

1.1. Parallel but Interactive Learning in Groups

When we are solving problems, we are unlikely to be the only ones solving them. Common goals, skill sets, and motivations among the members of a group entail that people will typically be surrounded by people solving similar problems to themselves. This is true for other animals as well. Finding food, mates, and protection are problems shared by animals within the same group, and copying of solutions is frequently observed across many species (Hurley & Chater, 2005; Sumpter, 2010). The situation of learning along with others who are searching for good solutions to the same problems has unique but general group dynamic patterns that make it an important topic of study. One commonly observed group-level pattern is convergence, by which the members of a group adopt more similar solutions with passing rounds of solution exchange (Nowak, Szamrej, & Latané, 1990). For example, when members of a group can see the music selections made by others in the group, the entire group selects more similar music than when the members are not informed of each others' selections (Salganik, Dodds, & Watts, 2006; Salganik & Watts, 2009). A second pattern is that when people have only access to the solutions of their immediate neighbors, then spatially determined clusters of similar solutions arise (Latané & L'Herrou, 1996). A single region from within a larger group will often show substantial consensus in its members' solutions, but different regions may show striking diversity. These patterns of convergence and clustering, as well as others, will be explored in the experiments to be described.

The experiments described in this chapter focus on interactive, parallel problem-solving situations. By "parallel," we mean that each individual in the group is providing complete solutions to a problem, and that their rewards are based only on the quality of their own solutions. In other situations, the

members of a group coordinate such that the entire group generates a single solution to a problem (Kearns, Suri, & Montfort, 2006; Roberts & Goldstone, 2011). Both situations have real-world counterparts. Parallel problem solving is, perhaps, the more common situation, because it is implicated whenever individuals are self-interested and it is in their self-interest to imitate one another.

The “interactive” in “interactive, parallel problem solving” refers to the influence that problem solvers have on one another via the public nature of their solutions. For animals, the intentional signals or unintentional cues left by others can be used to find food and shelter (Sumpter, 2010). For companies, solutions are made publicly available when they are instantiated in commercially available products. For lobster harvesters in Maine, solutions to the problem of where to harvest to maximize one’s intake of lobsters are publicly available because of the presence of physical lobster traps (Acheson, 2003). For scientists, solutions are published in scholarly journals and presented at conferences, at which point the solutions may influence other scientists. A striking example of this last phenomenon is that estimates of physical constants in science tend to be inaccurate during early attempts to measure them. Subsequent attempts to measure the constants become more accurate, but they also tend to deviate systematically from the correct value in the direction of the earlier measurements (Henrion & Fischhoff, 1986). That is, new estimates of a physical constant tend to be distorted toward previous estimates. Historically, this pattern has been observed for the speed of light, Planck’s constant, the charge of an electron, the mass of an electron, and Avogadro’s number. In discussing systematic deviations in estimates of the charge of an electron, Feynman, Leighton, and Hutchings (1997) write, “Millikan measured the charge on an electron by an experiment with falling oil drops, and got an answer which we now know not to be quite right. It’s a little bit off because he had the incorrect value for the viscosity of air. It’s interesting to look at the history of measurements of the charge of an electron, after Millikan. If you plot them as a function of time, you find that one is a little bit bigger than Millikan’s, and the next one’s a little bit bigger than that, and the next one’s a little bit bigger than that, until finally they settle down to a number which is higher.” The fact that estimates of physical constants can be demonstrated to be influenced by previous estimates is noteworthy because each estimate is, in principle, being estimated solely on the basis of an experiment. Even when we use scientific methods and controls to shield ourselves from being influenced by others’ solutions, we cannot resist being influenced.

We cannot help being influenced by others because, in most situations, it is not the best policy to resist this influence. Social psychologists have historically stressed situations in which peer influences—from tacit learning to overt conformity—lead to impaired creativity (Kerr & Tindale, 2004), distorted judgments (Asch, 1956), or even dysfunctional actions (Milgram, 1974). However, in most cases, taking advantage of what others have discovered is a smart strategy. Imitating others' solutions is useful when people in a group tend to face similar challenges, when it is costly to explore a problem space on one's own, when the environment changes relatively slowly so that what was useful for one person will still probably be useful for another person, and when individual uncertainty is high (Boyd & Richerson, 1985, 2005).

1.2. Bridging between Individual and Group Levels of Explanation

We are interested in the consequences for the group when individuals learn to solve problems and know about each others' solutions. Accordingly, we do not follow the standard method used in social psychology of testing one participant in the company of experimenters' confederates who are scripted to respond in particular ways (Asch, 1956). Asch's method is well justified from the perspective of creating a well-controlled experimental environment for exploring factors affecting individual choices to imitate. However, the cost of constraining the judgments of all but one participant in a group is that the group dynamics of imitation cannot be revealed. The impact of individual imitation choices on the group's performance can best be discovered by allowing all participants in a decision-making task to be naturally and spontaneously influenced by one another. Understanding the group dynamics of imitation and innovation is one of the main goals of our study, and so we give all group members the opportunity to influence, and be influenced by, each other.

One result of our decision to let every group member influence every other group member is that the proper unit for our statistical analyses will be the group rather than the individual. Rather than trying to eliminate dependencies between individuals, we allow dependencies but then treat the entire set of interdependent components (e.g. participants in one experimental session) as the unit of analysis. This choice is based on a theoretical commitment that coherent group of people is often a highly useful level, even explanatorily indispensable, level of description (Goldstone & Ashpole, 2004; Theiner, Allen, & Goldstone, 2010). Understanding collective behavior requires theoretical constructs above the level of the individual. One

of the primary motivations for many agent-based models is to provide a theoretical bridge across different levels of description. Consider Schelling's (1971) classic "simulation studies" of segregation. Schelling created agents belonging to two classes (represented by dimes and pennies) that are reasonably tolerant of diversity and only move when they find themselves in a clear minority within their neighborhood, following a rule like "If fewer than 30% of my neighbors belong to my class, then I will move." Despite this overall tolerance, the agents still divide themselves into sharply segregated groups after a short time. What is surprising is that this occurs even though no individual in the system is motivated to live in such a highly segregated world. Although hardly a realistic model of migration, the model has been influential in contrasting group-level results (i.e. widespread segregation) and individual goals. If group-level constructs like segregation, wealth disparity, monetary flow, social network topology (Kennedy, 2009), and intellectual climate are eliminated, then many of the most surprising and useful theoretical claims for how individual-level incentives affect these constructs would no longer be possible. Not only would we miss out on truly bridging theories that show how individual behavior creates behaviors at a completely different level, but we would also lose much of our ability to predict and control social structures at scales that are meaningful for society.

Applying this moral to our experiment on social learning in groups, we will be explicitly interested in creating bridging explanation between explanations at the individual and group levels. One of our primary interests is in the consequences for the group as a whole when individuals engage in individual versus social learning. Many of the properties we measure at the group level are not even meaningful constructs at the individual level. These properties include the collective coverage of a problem space by the group, the diversity of solutions within a group, and the prevalence of reciprocal copying in which A copies B's solutions, tweaks them, and then B copies A back. The existence of these quantifiable properties at the group, but not individual, level helps to warrant the belief that multiple levels of organization must be posited for explanatory and predictive validity.

1.3. Exploration and Exploitation

One of the most important bridges between individual and group behaviors concerns individuals' chosen positions along an exploration–exploitation trade-off (Hills, Todd, & Goldstone, 2010; Roberts & Goldstone, 2006). Exploratory behavior introduces new solutions by searching in hitherto unknown regions of a problem space. It tends to be high risk because

of the uncertainty about payoffs in unknown regions (Boyd & Richerson, 2005), but engaging in exploration can also have favorable long-term payoffs if the agent can take advantage of discovered bountiful resources for a prolonged period after the initial exploration (Sang, Todd, & Goldstone, 2011). Exploitation behavior involves taking advantage of solutions previously found, either by oneself or others. Rather than viewing exploration and exploitation as opposed to one another, they should be seen as reinforcing. The value of exploration is amplified exactly because the fruits of exploration are subsequently exploitable. In situations where there are few opportunities for subsequent exploitation, exploration is rarely a sound strategy. If there is only one chance remaining to harvest resources, exploration is usually a poor choice because there will not be any future opportunities to exploit what has been found. Exploitation is what makes exploration valuable.

Individual decisions to explore or exploit have powerful influences on the group's performance, and not always in a straightforward fashion. Exploiting the solutions of others through imitation is useful to the group because it allows effective innovations to spread. However, it can also reduce the group's overall ability to fully cover the range of potential solutions or options. As an example, of this reduced potential, Salganik and Watts (2009; Salganik, et al. 2006) allowed participants to download music from a site, sometimes with knowledge about the downloads made by their peers. By assembling participants into independent groups, they were able to measure whether separate "re-runnings of history" would have produced the same most popular songs, or whether different songs would arise as most popular because of rich-get-richer dynamics operating on initially haphazard choices. In fact, when participants had information about each others' download choices, then relatively imbalanced patterns of downloading arose, and some songs were downloaded far more often than others. For different groups, very different sets of songs became popular. The inequitable pattern of downloads compromised the groups' ability to collectively sample the full range of possible music. This is a classic example of choice copying reducing group performance by restricting the injection of new options. Other research has shown that early decision makers can have an undue influence on the group's behavior when subsequent decision makers are influenced by their own judgments as well as their predecessors' judgments (Bikhchandani, Hirshleifer, & Welch, 1992). Bettencourt (2009) formally models the importance of having sufficient independence among judges if the benefits of synergistic aggregation are to be achieved.

Individual decisions to explore bring in their own hazards for group performance. Exploration does inject new innovations from which the group can subsequently choose. However, the innovations come at the cost of underutilization and transmission of good solutions already at hand. If all members of a community are continually exploring new possibilities rather than taking advantage of existing solutions, then previous generations' solutions may be practically forgotten by newer generations. In the extreme, exploration without exploitation can halt the "cultural ratchet" that has been implicated in humans' unique ability to create lasting and improving cultural products (Tomasello, Kruger, & Ratner, 1993). This risk is not merely theoretical. Researchers have documented the collective forgetting of knowledge that would be useful for a community, such as an understanding of complex interactions among biological species in an ecosystem (Wolff, Medin, & Pankratz, 1999). Specific cultures, such as the Itza' Maya of Guatemala, have acquired over centuries knowledge of their natural world that is rapidly being left behind despite its continued relevance (Atran, Medin, & Ross, 2004).

Given the tradeoffs and interactions between exploration and exploitation, there will be no general solutions to the question of what percentage of one's time should be spent exploring versus exploiting. The answer to this question will depend on one's social orientation (whether one is seeking an optimal individual or group outcome), how many opportunities to seek solutions still remain, the complexity of the problem space, the density of one's social network, and the decisions that others are making to explore versus exploit.



2. INNOVATION PROPAGATION IN A ONE-DIMENSIONAL PROBLEM SPACE

In social psychology, there has been a long and robust literature on conformity in groups (Cialdini & Goldstein, 2004; Sherif, 1935). The usual finding is that people conform to majorities in groups. To some degree, conformity is found because people desire to obtain social approval from others. For example, sometimes when people give their answers privately, they are less likely to conform to the group's opinion than when responding publicly (Deutsch & Gerard, 1955). However, at other times, the conformity runs deeper than this, and people continue to conform to the group's opinion even privately (Sherif, 1935). In our experiments and modeling, we are interested in the use of information provided by others even when social

approval motivations are minimized because the group members never meet one another and are anonymous.

Conformity to others' ideas has been a major field of research not only in social psychology but also in economics, political science, and sociology. It is common in models of collective action to make an individual's decision to participate based upon their expectations for how many other people will participate (Chwe, 1999). A common outcome of a collective "I'll do it if you do it" mentality is for "tipping points" to arise in which adding more participants to an action leads to a positive feedback cycle in which still more participants sign on, leading to an exponential increase in participation for a time (Gladwell, 2000). This behavior is a sensible policy both because the likelihood of success of an innovation depends upon its public adoption rate (Bullheimer, Dawid, & Zeller, 1998) and because other people may have privileged information unavailable to the individual making a choice. The potential cost of this bandwagon behavior is wasted time, money, and effort in adopting new innovations when existing solutions are as good or better. Furthermore, bandwagons entail redundant convergence on a single solution rather than continued broad search of a problem space (Rosenkopf & Abrahamson, 1999; Strang & Macy, 2001).

Our studies explore the diffusion of innovative ideas among a group of participants, each of whom is trying to individually find the best solution that they can to a search problem. The work fills an important gap in research. There are several promising computational models for how agents in a population exchange information (Axelrod, 1997; Kennedy & Eberhart, 2001; Nowak *et al.*, 1990). There is also excellent work in social psychology on how individuals conform or use information provided by others (Gigone & Hastie, 1996). Fieldwork also explores actual small groups of people engaged in cooperative problem solving (Arrow, McGrath, & Berdahl, 2000). However, there is very little work with laboratory-controlled conditions that explores the dynamics of a group of participants solving problems as they exchange information. One related study is Latané and L'Herrou's (1996) exploration of participants' sending e-mail messages to each other (Latané & Bourgeois, 1996) as they tried to predict which of two options their group would select. Over the course of message exchanges, neighboring participants in the network tended to adopt similar choices (consolidation), but there was also continued diversity of choices across the entire network. In contrast to this work, our research predominantly focuses on situations where participants are trying to find good solutions to a problem rather than trying to conform to their neighbors.

For example, farmers may discuss the benefits of various crop rotation techniques with their neighbors, and may be convinced to try a new one by a neighbor's success, but there is no reward to conforming to a neighbor's behavior in itself. Other research in this area has recently appeared (Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Mason & Watts, 2011), and is likely to expand, given its relevance to parallel, independent collective search processes in businesses, the internet, and elsewhere.

2.1. An Experimental Examination of Connectedness and Fitness Functions

In creating an experimental paradigm for studying information dissemination, our desiderata were (1) a problem to solve with answers that vary continuously on a quantitative measure of quality, (2) a problem search space that is sufficiently large that no individual can cover it all in a reasonable amount of time, and (3) simple communications between participants that are amenable to computational modeling. We settled on a minimal search task in which participants guess numbers between 0 and 100 and the computer reveals to them how many points were obtained from the guess by consulting a hidden fitness function (Mason, Jones, & Goldstone, 2008). Additionally, random noise was added to the points earned, so that repeated sampling was necessary to accurately determine the underlying function relating guesses to scores. Over 15 rounds of guesses, participants try to maximize their earned points. Importantly, participants get feedback not only on how well their own guess fared but also on their neighbors' guesses. In this manner, participants can choose to imitate high-scoring guesses from their neighbors. We experimentally manipulated the network topology that determines who counts as neighbors, as well as the fitness function that converts guesses to earned points.

We created neighborhoods of participants according to random, regular lattice, fully connected, and small-world graphs. Examples of the graph topologies for groups of 10 participants are shown in Figure 1.1. In the random graph, connections are randomly created under the constraint that the resulting graph is connected—there is a path from every individual to every other individual. Random graphs have the property that individuals tend to be connected to other individuals via paths that do not require passing through many other individuals. This property has been popularized as the notion of “six degrees of separation” connecting any two people in the world, and has been experimentally supported (Dodds, Muhamad, & Watts, 2003; Milgram, 1967). More formally, the average path length connecting two

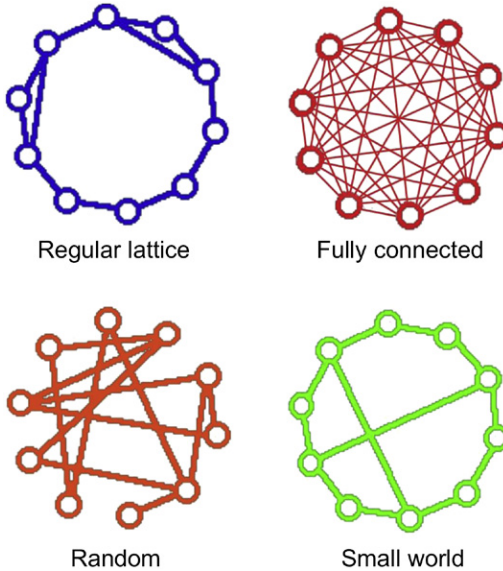


Figure 1.1 Examples of the different network structures for groups of 10 participants from the experiment on collective search in a one-dimensional problem space (Mason *et al.*, 2008). Circles represent participants and lines indicate communication channels. For color version of this figure, the reader is referred to the online version of this book.

randomly selected nodes in a random graph is $\ln(N)/\ln(K)$, where N is the number of nodes and K is the average number of neighbors connected to each node. The regular lattice can be used to represent a group with an inherent spatial ordering such that people are connected to each other if and only if they are close to one other. The regular lattice also captures the notion of social “cliques” in that if there is no short path from A to Z, then there will be no direct connection from any of A’s neighbors to any of Z’s neighbors. In regular lattices, the average path required to connect two individuals requires going through $N/2K$ other individuals. Thus, the paths connecting people are much longer, on average, for lattice than random graphs.

Random graphs have short paths, but unfortunately (from the perspective of realistic modeling of social phenomena) do not contain cliques. Lattices show cliques, but do not have short path lengths. Recently, considerable interest has been generated in networks that have both desirable properties, the so-called “small-world networks.” These networks can be formed by starting with a lattice and randomly rewiring (or adding new connections, in the case of our experiments and Figure 1.1) a small

number of connections (Watts & Strogatz, 1998). The result is a graph that still has cliques because nodes that are connected to the same node tend to be spatially close themselves, yet also have a short average path length. From an information processing perspective, these are attractive networks because the spatial structure of the networks allows information search to proceed systematically, and the short-cut paths allow the search to proceed quickly (Kleinberg, 2000). Notice, in Figure 1.1, that all three of the described networks have a total of 12 connections between 10 participants. Thus, if there is a difference in information dissemination in these networks, then it must be due to the topology, not density, of the connections. A fourth network, a fully connected graph, allowed every participant to see the guesses and outcomes of every other participant.

We compared two hidden functions for converting guessed numbers to points. The unimodal function has a single best solution that can always be eventually found with a hill-climbing method (Figure 1.2a). The trimodal function (Figure 1.2b) increased the difficulty of the search by introducing local maxima. A local maximum is a solution that is better than all of its immediate neighboring solutions, yet is not the best solution possible. Thus, a simple hill-climbing search might not find the best possible solution.

Twelve groups of Indiana University undergraduate students ranging in size from 7 to 18 people with a median of 14 people per group participated for partial course credit, for a total of 153 participants. Each group participated in eight experiments that consisted of every combination of the four network types (Figure 1.2) and two fitness functions (Figure 1.2). Participants were told to try to maximize their total number of points acquired over 15 rounds of number guessing, and that the same guess would be worth about the same number of points from round to round, but that a certain amount of randomness was added to the earned points. Participants were also told that they would see the guesses and points earned by some of the other participants, and that these others would also see the participants' guesses and earnings.

The results from this experiment are shown in Figure 1.3, expressed in terms of the percentage of participants within one-half standard deviation of the global maximum for a fitness function (similar results are found if "total points" is used as a dependent measure). Over the 15 rounds, increasingly many participants find the global maximum. For the unimodal function, the fully connected network finds the global maximum most quickly, and the advantage of the fully connected network over the other three networks is particularly striking for Rounds 2–4. Around Round 5, the

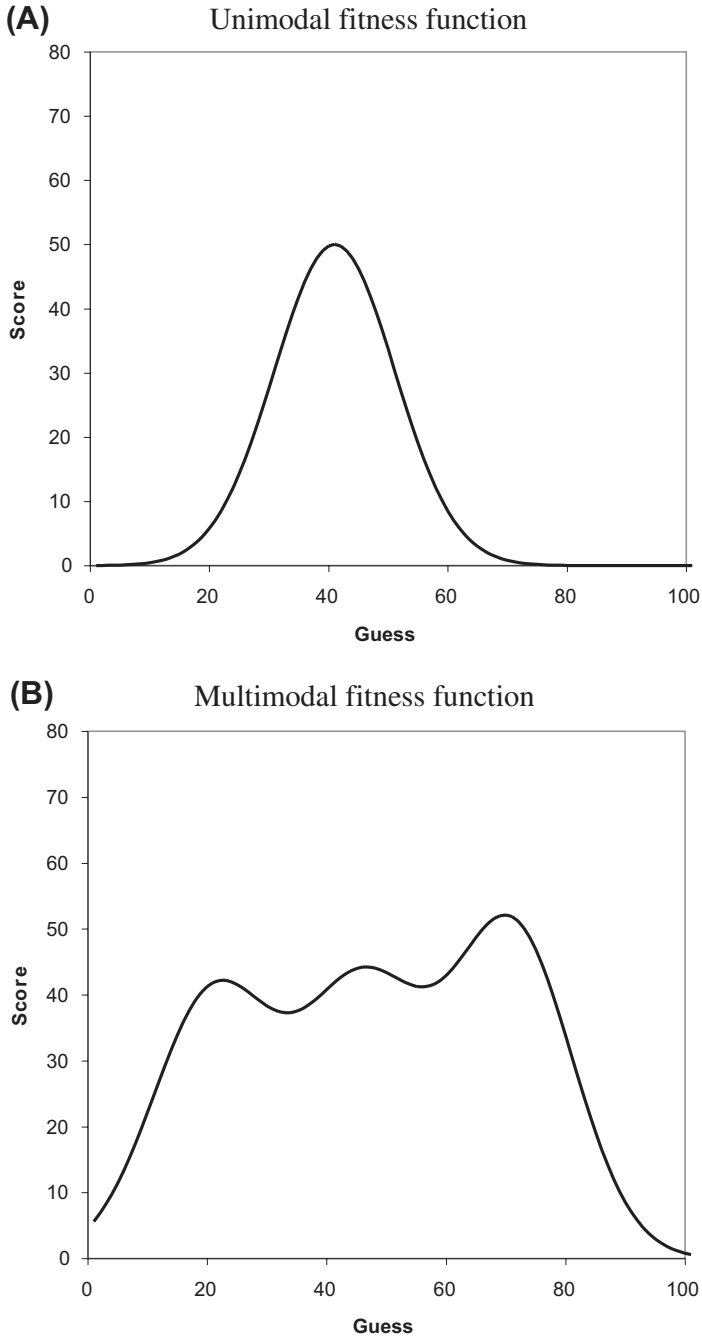


Figure 1.2 Examples of the unimodal and multimodal fitness functions that convert guesses into obtained points.

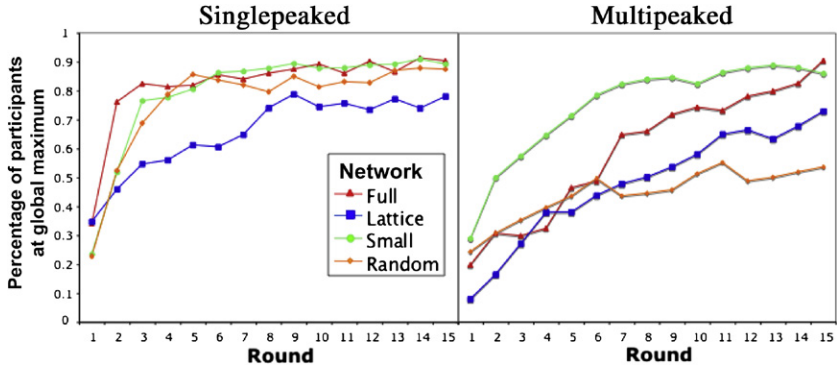


Figure 1.3 Percentage of participants within one standard deviation of the global maximum on each round for the unimodal and multimodal payout functions. For color version of this figure, the reader is referred to the online version of this book.

small-world network catches up to the performance level of the fully connected network, and for the rest of the rounds, these two network types continue to outperform the other two networks. This pattern of results is readily explainable in terms of the propensity of a network to disseminate innovations quickly. Innovations disseminate most quickly in the full network because every individual is informationally connected to every other individual.

For the multimodal payout function, the small-world network performs better than the fully connected network for the first six rounds. One account for its superiority over the full network is that the small-world network is able to thoroughly search the problem space. The fully connected groups frequently get stuck in local maxima because the groups prematurely converge on a good, but not great, solution. The small-world structure is an effective compromise between fully exploring a search space and also quickly disseminating good solutions once they are found. The most surprising aspect of these results is that the truism of “the more information, the better” is not supported. Giving each participant all of the results from all of the agents does not lead to the best group solution for the multimodal problem—the downside of this policy is that with the fully connected network, everybody ends up knowing the same information. Participants thereby become too like minded, acting like a single explorer, rather than a federation of independent explorers.

The general point from this first experiment is that before one decides how to connect a group, one should know about the nature of the problem the group needs to solve. A candidate generalization is that the more

exploration a group needs to do, the more clustered and locally connected the network should be. Conversely, the more quickly a group needs to exploit emerging solutions, the more globally connected individuals should be. Problem spaces that require considerable exploration to find the global maximum should benefit from networks that have relatively well-isolated neighborhoods that can explore different regions of a problem space. To test this hypothesis, in a separate experiment, we tested the more difficult fitness function shown in [Figure 1.4](#), which we call the needle function because of the thin and high global maximum and because finding this global maximum is a bit like finding a needle in a haystack. This function features one very broad local maximum, and one hard-to-find global maximum. We tested 12 groups of participants in needle functions like [Figure 1.4](#), with each group connected in the same four network topologies we used before. For this function, [Figure 1.5](#) shows that the lattice network performed better than the other three network types, starting by Round 7, if not earlier. The lattice network fosters the most exploration because of its spatially segregated

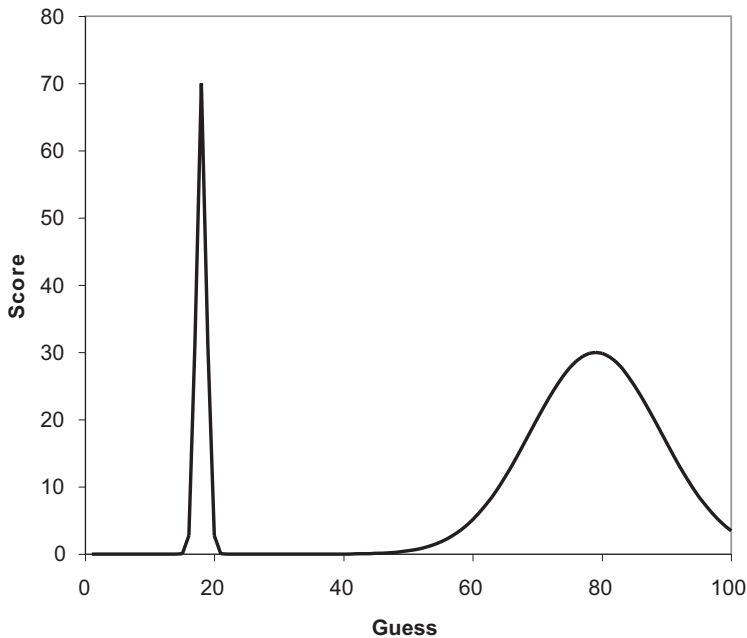


Figure 1.4 An example of the “needle” payout function. This function features one broad local maximum that is easy to find and one narrow global maximum that is difficult to find.

network neighborhoods. Exploration of the problem space is exactly what is needed for the needle function because of its hard-to-find global maximum.

The three payout functions are ordered by the demands they place on broad exploration of a problem space. The benefit for exploration increases going from the unimodal to the multimodal to the needle function. In parallel, the network structures are ordered by their preservation of local cliques of nodes. Cliquishness increases going from full to small world to lattice networks. These two progressions are coordinated, as is shown in Figure 1.6, with both progressions going from the left to the right. The full network performs best with the unimodal function, the small-world network performs best with the multimodal function, and the lattice performs best with the needle function. In contrast to arguments for a general informational advantage of small-world networks (Watts & Strogatz, 1998), we find that what network is best depends on the kind of problem a group

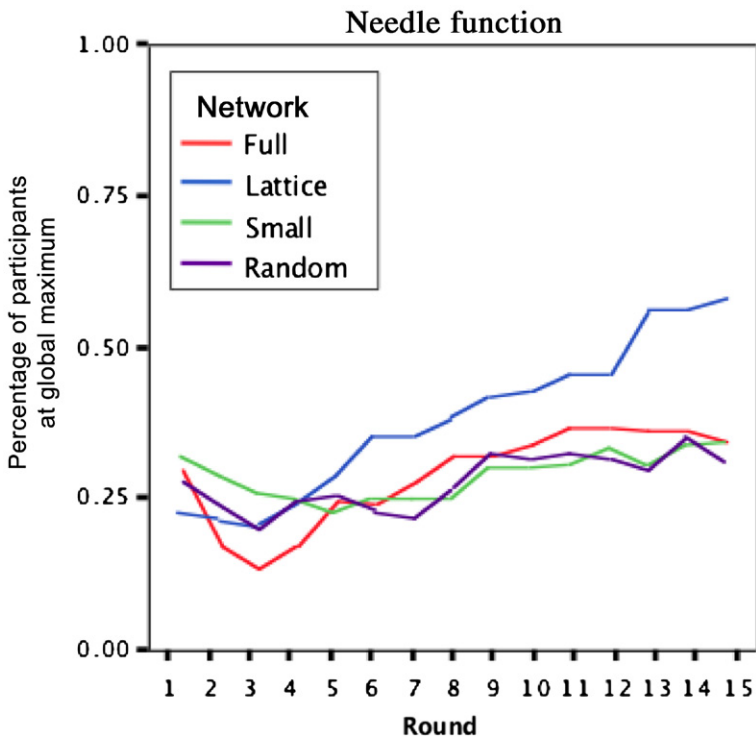


Figure 1.5 Performance for the four network structures with the needle payout function. For this function, the lattice network performs better than the other three network types. For color version of this figure, the reader is referred to the online version of this book.

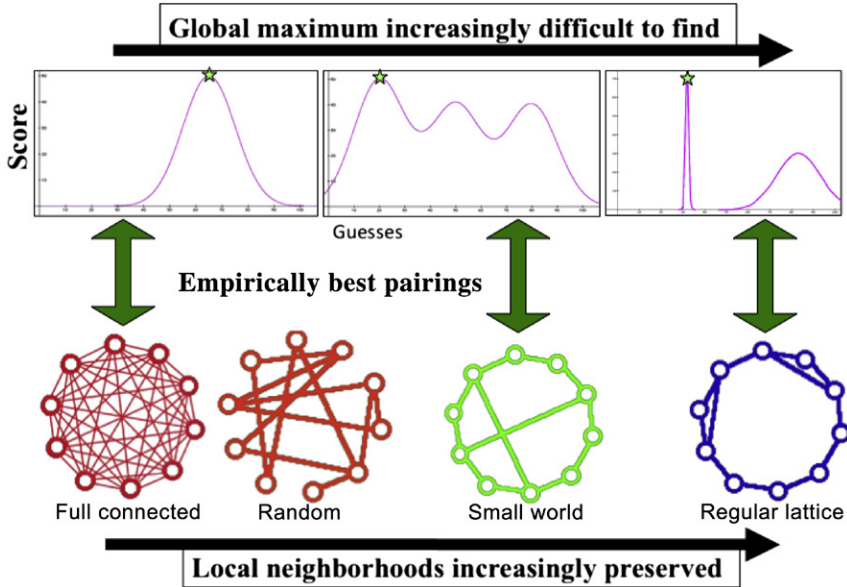


Figure 1.6 A summary of the empirical relationship between the difficulty of a problem search and the best-performing social network. As it becomes increasingly difficult to find the global maximum in a problem space (indicated by stars), the best-performing network has increasingly preserved local neighbors. Long-range connections are good for rapid diffusion of optimal solutions once they have been found, but risk premature convergence of the entire network onto good, but not optimal, solutions. For color version of this figure, the reader is referred to the online version of this book.

must solve (Lazer & Friedman, 2005). As broader exploration is needed to discover good solutions, increasingly cliquish networks are desirable.

2.2. A Computational Model of Innovation Propagation

We have developed an agent-based computational model of our experiments based on the premise that members of a group can choose to explore a problem space on their own or take advantage of the solutions found by others. In the model, called SSEC (for Self-, Social-, and Exploration-based Choices), every agent on every round probabilistically chooses between three strategies: using their own guess on the last round, using their neighbors' best guess on the last round, and randomly exploring. Each agent randomly chooses between these strategies, with the likelihood of each strategy based on its intrinsic bias and also its observed success. The model, thus, can be expressed as

$$p(C_x) = \frac{B_x S_x}{\sum_n B_n S_n},$$

where $p(C_x)$ is the probability of using Strategy x , B_x is the bias associated with the strategy, and S_x is the score obtained from the strategy. The participant's guess is then $G_x + N(\mu = 1, \sigma = 1)$, including normally distributed randomness to avoid perfect imitation, with G_x being the guess associated with Strategy x . When the random exploration strategy is selected, a uniform distribution is used to select the next guess. This model is motivated by the Particle Swarm Algorithm (Kennedy, Eberhart, & Shi, 2001). However, unlike the swarm algorithm, the SSEC model allows sudden jumps in guesses rather than smoothly changing patterns of oscillations around promising solutions. The experimental results showed that participants frequently jumped from one guess to a completely different guess, a behavior that the original Particle Swarm Algorithm does not accommodate.

The simplest version of this model, with mostly default parameter values for the biases, was able to accommodate some, if not all, of the trends in the results. In particular, we tested a version of the model in which B_1 (the bias for using one's own previous guess) is 1, B_2 (the bias for using one's neighbor's best-scoring guess) is 1, and B_3 (the bias for randomly exploring) is 0.1. This is essentially a one-parameter control of biases because B_1 and B_2 were constrained to be equal, and only the relative, not absolute value of B_3 matters given the choice model used to determine strategy choice. In addition, the value of σ that determines the mutation/drift rate for guesses was set to 3, and noise with a variance of 30 and a mean of 0 was added to the fitness function's output, just as it was to experimental scores. Each of the four network types was run 1000 times with each of the three fitness functions for 15 rounds of guessing and 15 agents per group. In this model, fully networked groups were best for the unimodal function, small-world groups were best for the small-world network, and latticed groups were best on the needle function. The model predictions are shown in Figure 1.7, and can be compared to the human results in Figure 1.3. The fit is not perfect, but even with no parameters optimized for fit to the human data, roughly similar trends are found for both the model and humans.

Given the promising results of this original set of simulations, we parametrically manipulated the network connectivity to continuously shift from a regular lattice with only local connectivity to a fully connected network in which every agent is directly connected to every other agent. This was achieved by connecting 15 agents via a lattice, and then adding a number of additional random connections between agents. As the number of random connections increases, the network initially transforms from a random network to a small-world network. Then, as the connectivity further increases,

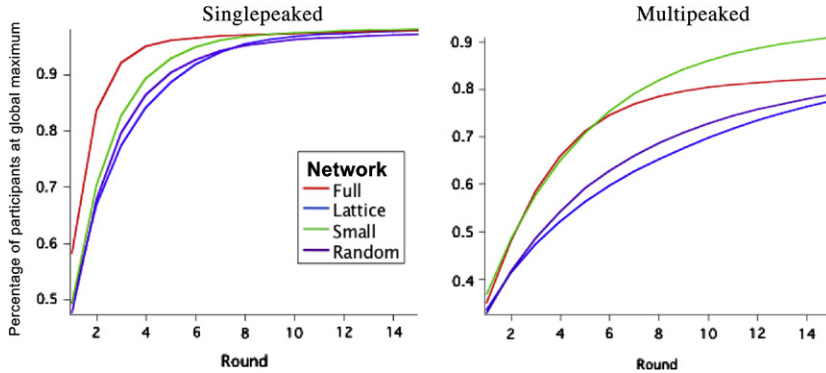


Figure 1.7 Simulations results for the SSEC model when given the single-peaked and multip peaked fitness functions. For color version of this figure, the reader is referred to the online version of this book.

the network transforms from a small-world network to a fully connected network. If more information communicated in a network always increases group performance, then we expect better performance (shown by brightness in Figures 1.8–1.10) as connectivity increases.

Independently, we manipulated the relative weight given to information obtained from oneself compared to others. Keeping B_3 constant at 0.1, we varied B_1 from 0 to 1 and set B_2 equal to $(1-B_1)$. Thus, we varied the degree to which each agent’s guesses are based on their own previous guess compared to others’ guesses. In Figures 1.8–1.10, as we go from the left to the right, we go from “sheepish” agents that base their guesses completely on others’ guesses (and an occasional random guess) to “mavericks” that always continue using their own solutions without any influence of others.

Figures 1.8–1.10 show that the influences of connectivity and agent independence are not constant, but rather depend on the shape of the problem space. For the easy-to-solve unimodal problem, Figure 1.8 shows that group performance increases monotonically with both increased reliance on others’ information and increased connectivity. Both trends can be explained by the fast propagation of innovations obtained when agents follow their best peers, and have many peers to follow. For single-peaked problems, there are no local maxima and so no concern with hasty collective convergence on suboptimal solutions.

For the three-peaked function (Figure 1.9), optimal group performance involves intermediate levels of both connectivity and self-reliance. These two factors trade-off with one another such that increases in connectivity can be offset by decreases in conformity. Networks that have only local

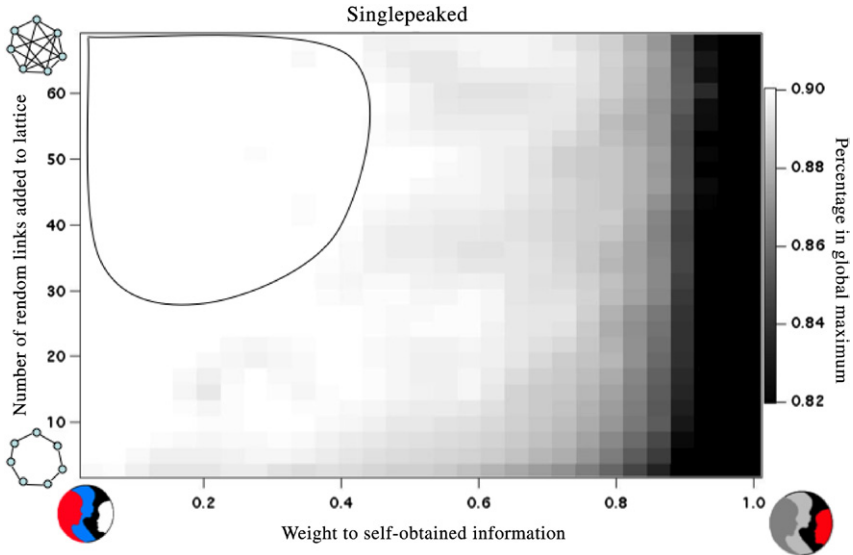


Figure 1.8 Group performance for the single-peaked function. This graph shows the interaction between the bias for self- versus other-obtained information and the number of random links added to a regular lattice. Group performance is measured by the percentage of individuals within one standard deviation of the global maximum of the fitness function. The brightness of each square indicates the group’s performance after 15 rounds of number guessing. The area of the parameter space that produces the best performance is outlined in black. For this simple problem space, group performance increases monotonically with increased reliance on others’ information and network connectivity. For color version of this figure, the reader is referred to the online version of this book.

connectivity and self-reliant individuals perform relatively poorly because good solutions are inefficiently spread. Conversely, networks that have global connectivity and conformist individuals also perform poorly because the group frequently converges on local rather than global maxima. Good group performance is found when a group can both search a problem space for good solutions, and yet spread those solutions quickly once they are found. This is achieved when conformist individuals communicate over a sparsely connected network, or when self-reliant individuals communicate over a more broadly connected network. If one is able to engineer a social network, then one’s target network should depend on both the problem and “personalities” (mavericks vs. sheep) of the nodes in the network.

For the trickier needle function (Figure 1.10), the best-performing networks are pushed even further in the direction of increasing self-reliance and decreasing connectivity. Consistent with our empirical results, the

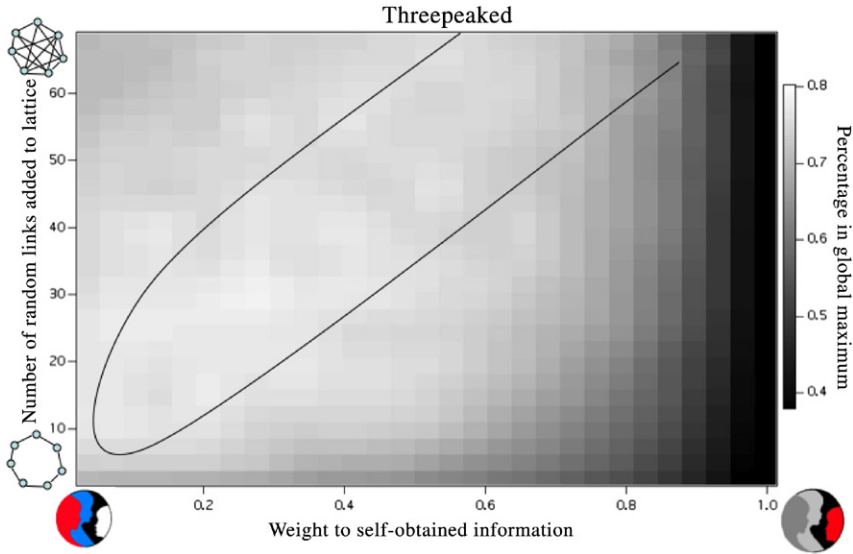


Figure 1.9 Group performance for the multimodal function. The best performance is found for a combination of using self- and other-obtained information, and for intermediate levels of network connectivity. As the degree of connectivity increases, best group performance is achieved by decreasing reliance on others' guesses. For color version of this figure, the reader is referred to the online version of this book.

needle function requires more exploration, and both limiting connectivity and increasing self-reliance promote independent exploration of group members. As with the three-peaked function, there is a trade-off between network connectivity and individual self-reliance.

A major conclusion from both the experiments and modeling is that propagating more information is not always good for the group. Full access to what everybody else in a group is doing can lead human and computational agents to prematurely converge on suboptimal local maxima (Lazer & Friedman, 2005). Networks that preserve spatial neighborhoods promote exploration, and this can explain why the full network is the best network for the single-peaked function, the small-world network and its intermediate level of connectivity does best with the three-peaked function, and the lattice function with no long-range connections does best with the difficult needle function.

Although more information is not always better as far as the group goes, it is always in the best interest of individuals to use all the information at their disposal. Accordingly, our innovation propagation paradigm provides an unexpected example of a social dilemma (Goldstone & Janssen, 2005; Ostrom *et al.*, 1994). Individuals, looking out for their own self-interest, will

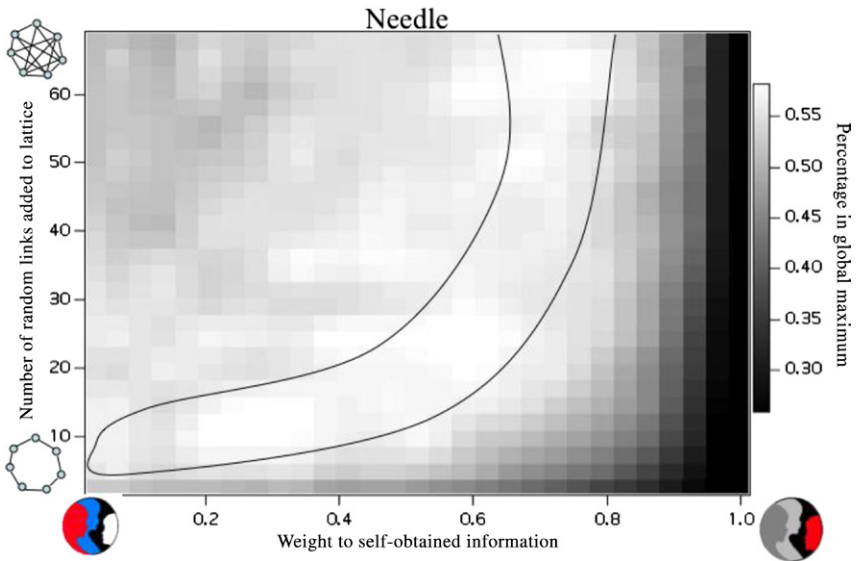


Figure 1.10 Group performance for the needle function. This function benefits from even greater reliance on self-obtained information and decreased global network connectivity, as shown by rightward and downward displacement of the best-performing ridge of the parameter space. For color version of this figure, the reader is referred to the online version of this book.

seek out as much information from others as possible, but this can inhibit the group as a whole from widely exploring a search space. Thus, in the present situation, obtaining information from as many peers as possible is noncooperative behavior even though it involves conformity. Searching a problem space on one's own is cooperative in the sense of allowing the group as a whole to collect the most points possible, by avoiding local maxima. Our simulations show that every individual agent is best off linking to as many other people as possible. Agents with relatively many links outperform those with relatively few links. However, if every agent links maximally to every other agent, then the entire group does not perform well due to premature convergence on good, but not optimal, solutions. Sensitivity to this conflict between individual and group interests may help in the design of adaptive social networks. Designing for the greater common good may sometimes entail placing limits on individuals' ability to connect with each other. Problems with difficult, hard-to-find solutions often drive people to look to others for hints and clues, but these are exactly the kinds of problems for which limited, local connectivity is advantageous.

This analysis of the conflict between the good of the individual and group becomes particularly relevant when we turn to situations where people can choose their connectivity, rather than having it imposed. Pursuing experimental paradigms in which people can create their own social networks would be valuable as connecting with both the mathematical literature on the evolution of networks (Dorogovtsev & Mendes, 2003) and the social science literature on coalition formation (Kahan & Rapoport, 1984). In many naturally occurring groups, people have some choice in who they will share information with, and what information they will reveal. From our perspective on human groups as complex systems, one of the interesting issues will be to study the global efficiency of information transmission in self-organized networks, and how incentives to individuals can be structured so that globally advantageous networks emerge.



3. COLLECTIVE LEARNING IN HIGHER-DIMENSIONAL PROBLEM SPACES

One dissatisfaction with the initial experimental paradigm is that the problem space is not particularly complex. For some fitness functions in the first experiment, it was difficult for participants to discover the global maximum, but this was due to the limited number of guessing rounds and the narrow basin of attraction for the global maximum. The second experiment was designed to provide a better experimental analog to a collective search situation in which members of a community are generating novel innovations to a relatively open-ended problem. Scientists coming up with new experimental paradigms, sports teams coming up with new plays, and artists coming up with new styles are all engaged in a search for innovations with a problem space that is impossible for a single individual to cover by themselves over a realistic time period. We chose an experimental paradigm most closely resembling the last of these situations, in which participants see drawings created by others as they create their own (Wisdom & Goldstone, 2011). Unlike a community of artists, we incorporated a simple, objective measure of the quality of drawings, so that we could inform participants of the quality of each others' solution. Unlike the first experiment, we only incorporated a fully connected network in which every participant could see every other solution on every round. The fully connected network seems like a natural, minimally assumptive default network, and offers the greatest potential influence of others on one's own decisions.

Using this paradigm, we were interested in describing individuals' strategies for imitating and innovating, and the consequences of these strategies for the group as a whole. For example, in the relatively constrained problem space of the first experiment, we found a tension between individual and group outcomes, with imitation being good for the individual but bad for the group. If this is replicated in the current experiment, it will suggest some generality to the social dilemma of innovation. If not, it will suggest a relation between the nature of a problem space and the existence of social dilemmas.

More generally, the collective drawing task allows us to observe participants' strategies for innovating and imitating. There are possible strategies related both to *which* participants' drawing to imitate and *when* to imitate (Laland, 2004). We might expect for participants to imitate other drawings that are scoring well, and that are scoring better than their own drawings (Rendell et al., 2010). It is also possible that drawings will be imitated that are already relatively similar to an imitator's current drawing, if the imitator finds it difficult or risky to blend potentially incompatible solutions. Participants might be expected to imitate more at the beginning of a set of rounds, when their uncertainty is the greatest, when their scores are relatively poor, and when there is a diverse range of possible solutions (Galef & Laland, 2005).

3.1. The "Draw the Mystery Picture" Task

With these predictions in mind, 145 participants were distributed into 39 groups ranging in sizes from 1 to 9. The participants' task was a round-based picture-matching puzzle game with score feedback given after each round. The goal picture that participants attempted to match was a randomly generated spline quantized to a grid of square pixels. The participants' game board was a grid of the same dimensions as the goal picture, with each square initially colored white. The color of each square on the game board could be toggled between black and white by clicking it with the mouse. Each participant's display included their own game board and the most recent score (given as the number of squares, both black and white, marked correctly out of the total number of squares on the board), their neighbors' game boards and scores, and indications of the current round in the game and the amount of time remaining in the current round (Figure 1.11). Players could copy a neighbor's most recent solution to their own at any time during the game by clicking the chosen neighbor's board with the mouse. Each game consisted of 24 rounds of 10 s each. After the last round in each

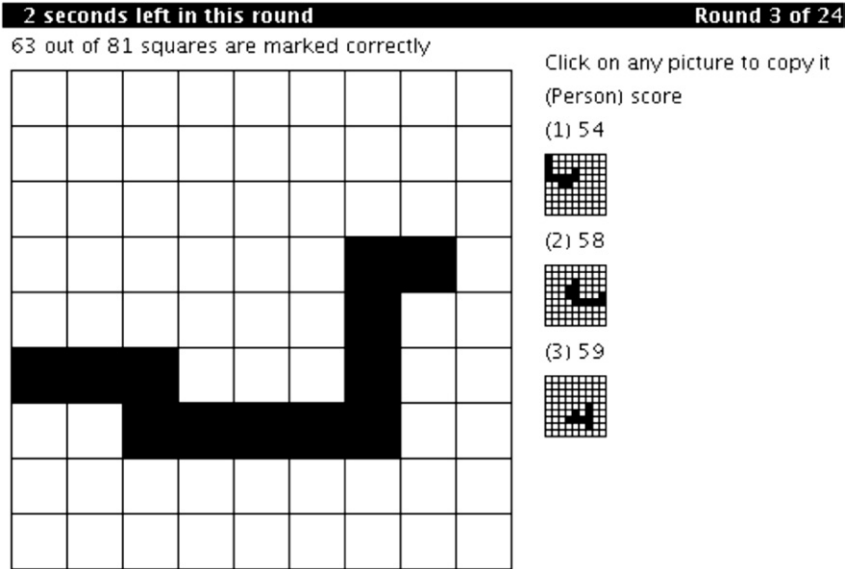


Figure 1.11 Screenshot of a participant's display in experiment 2. The participant's own created picture is the large image to the left. All other participants' pictures from the previous round, and the scores earned by these pictures, are shown to the right.

game, participants were shown their guesses and scores for each round, along with the goal picture, and a button to click when they were ready to begin the next condition. Participants were instructed to maximize their scores over all rounds by matching the hidden goal picture as closely as possible.

A participant's *score* in each round was defined as a cell-by-cell comparison (overlap) between the participant's guess for that round and the hidden goal picture (i.e. the number of cells which the two pictures had in common), divided by the total number of squares in the goal picture, to give a percentage that could be compared between conditions of varying grid size (Figure 1.12). This same overlap measure was used to determine the similarity between two different drawings. An *improvement* was defined as an instance of a participant obtaining a score higher than all prior scores of all players within a particular condition. *Turnover* for each round (after the first) was a measure of the amount of change between a participant's guesses over successive rounds. It was defined conversely to similarity, except that the two pictures compared were the participant's guesses from the current and previous round. A participant was regarded to be imitating another participant in a particular round if the participant's guess was closer to the most similar neighbor's previous guess than to the participant's own previous

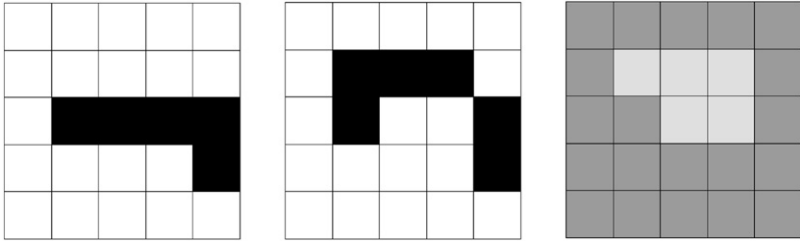


Figure 1.12 Determination of the similarity of two solutions. The left and middle pictures have the same black/white state for 20 out of 25 squares, shown in dark gray on the rightmost image, and so they have a similarity of 80%.

guess. *Diversity* (a measure of the spread of group members' guesses over the problem space within a particular round) was defined as follows:

$$D_r = 1 - \frac{\sum_s \sum_p \text{majority}(G_{\text{spr}})}{S_{\text{tot}} P_{\text{tot}}},$$

where G_{spr} is the binary value (black or white) of square s in the guess of participant p in round r , S_{tot} is the total number of squares in the game board, P_{tot} is the total number of participants in the group, and *majority* is a binary function that conveys whether the value of G_{spr} is in agreement with the majority of participants in the group for that square in that round (0 = not in majority, 1 = in majority). Diversity as defined above is constrained to be within the 0–1 range, and higher values of diversity indicate more deviation of individuals' guesses from the majority guesses.

3.2. Major Results and Implications

Overall, the average guess turnover rate per round was 7.3% of the game board, and participants engaged in imitation on 25.8% of guesses. In the aggregate, participants achieved final scores that had 89.3% agreement with the best score. Scores reliably improved with passing rounds. Turnover rate, guess diversity, and imitation rate all decreased with passing rounds, as participants converged on better drawings. All these effects are shown in [Figure 1.13](#). In addition to these effects, increasing group size was associated with higher individual performance as well as higher imitation rate, presumably because more peers offered participants more options for imitation. Nearly all instances of imitation were of those with higher scores than the imitator's, implying that, like other animal species ([Templeton & Giraldeau, 1996](#)), people are biased toward

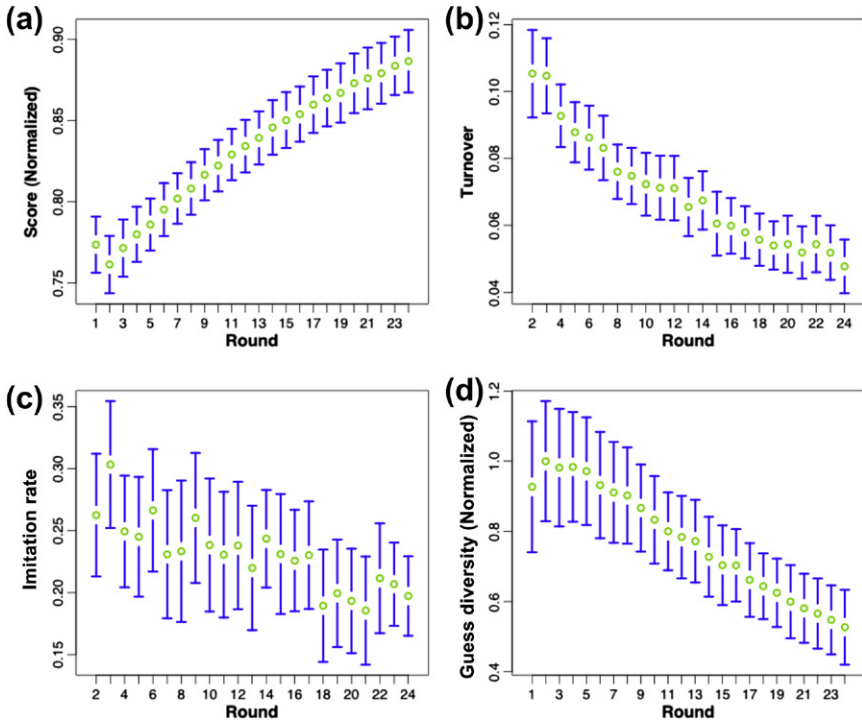


Figure 1.13 Results from experiment 2: (a) mean score increased, while (b) turnover, (c) imitation rate, and (d) guess diversity decreased as more rounds were played within a game. For color version of this figure, the reader is referred to the online version of this book.

imitating better-performing peers. The bias toward imitating the best-scoring peer was sizeable for small groups, but less pronounced for large groups, probably because participants were informationally overloaded by too many options.

To further investigate the relationship between strategy and performance, we performed regression analyses of score versus mean rates of imitation and turnover for individuals and groups. A linear regression of mean individual score versus mean individual imitation rate showed a significant positive relationship for individuals in group sizes of ≤ 4 , but none in groups of ≥ 5 (Figure 1.14a). Figure 1.14b shows that across all group sizes, there was a significant positive relationship between an individual's score and the mean imitation rate of all other group members, excluding the individual. That is, regardless of what an individual did, she/he was likely to have a higher score if the others in her/his group imitated more often. Figure 1.14c and d show a strong negative relationship between score and mean turnover. As one's

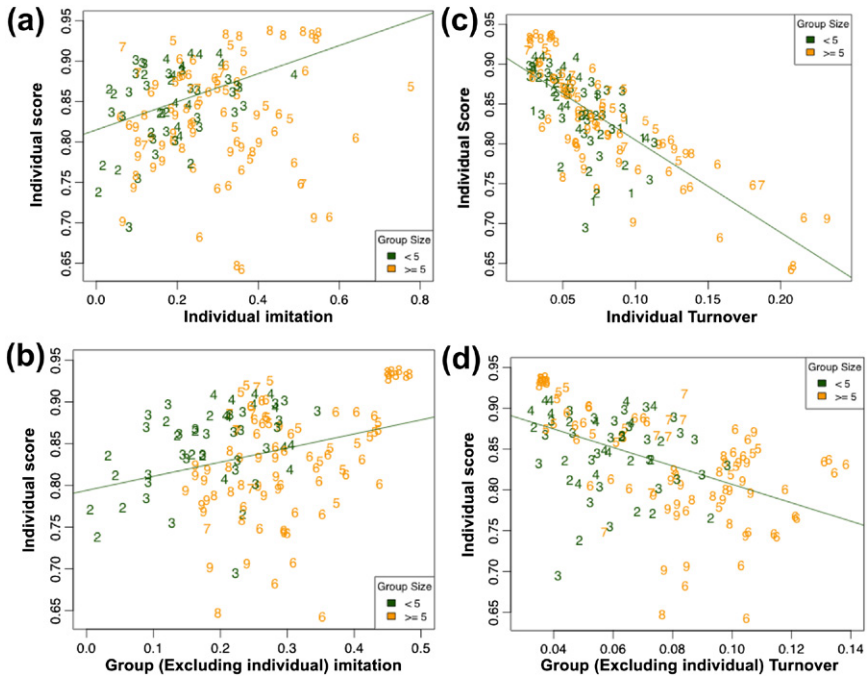


Figure 1.14 Results from experiment 2: (a) for smaller groups (<5 participants), higher imitation rates led to higher scores; however, these relationships did not hold for larger groups. (b) For all group sizes, regardless of a particular individual's imitation rate, the individual's score tended to increase as the imitation rate of others in the group increased. (c, d) Higher scores were associated with lower turnover rates. For color version of this figure, the reader is referred to the online version of this book.

peers turnover their guesses more, one's own score tends to be lower. The results from Figure 1.14c and d stand in striking contrast to the results from the simpler search space of the first experiment, in which adding opportunities for imitation (by letting every participant see every other participant) increased individual performance but decreased group performance.

Figure 1.15 shows a comparison between the similarity of imitators' most recent guesses to those which they imitated, and to those which they did not imitate. The analysis revealed that there was significantly greater similarity to imitated guesses than to nonimitated guesses (77.7% for imitated vs. 72.3% for nonimitated). In other words, imitation tended to be biased toward guesses that were more similar to the imitator's own prior guess. This difference held over all rounds within a game (Figure 1.15b), even though mean guess diversity decreased such that solutions generally converged (Figure 1.13d).

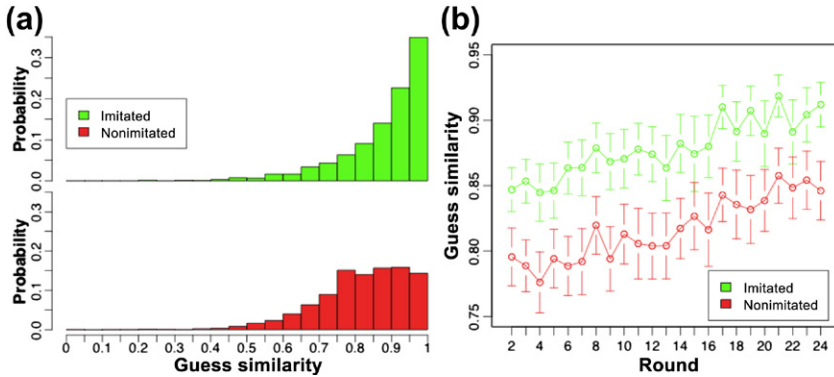


Figure 1.15 Similarity bias for imitation. (a) The probability of an imitation/nonimitation event occurring, as a function of the similarity between a participant's and peers' solutions. Imitators' previous guesses showed greater similarity to the guesses they imitated than to those they did not imitate. (b) The bias toward imitating more similar guesses was consistent across all rounds in a game. For color version of this figure, the reader is referred to the online version of this book.

Overall, participants' solutions improved over rounds through the use of fairly conservative strategies, as evidenced by the low mean turnover rate. Rather than large, revolutionary changes, participants made small, incremental improvements by changing only a few cells, typically just one. These small changes allowed participants to make accurate comparative inferences about their effects on score. Participants' rates of imitation and general turnover decreased across rounds, and the imitation that did occur was biased toward more similar solutions. This entrenchment of solutions carried over to the group level as well, shown by the decreasing group solution diversity across rounds.

The association of higher scores with greater imitation rates at both the individual and group levels shows that imitation is not always harmful to innovation and performance improvements. The rate of imitation was about the same among solutions that were improvements and nonimprovements, suggesting that improvements were often achieved by imitating a relatively successful participant's solution and then slightly tweaking this solution. Once tweaked, the improved solution was then available to other participants, including the individual who was originally imitated. The association of high individual scores with high imitation rates by others in the group (regardless of the individual's behavior) reinforces the idea of a systemic benefit for imitation rather than a view of imitation as a purely self-benefiting act. It may be that, regardless of the intentions of individuals, imitation benefits the

group by acting as a filter for propagating and preserving the better solutions available in a group at a given time, as was found in a recent competition of social learning strategies in a simulated environment (Rendell et al., 2010).



4. COLLECTIVE SEARCH IN A PROBLEM SPACE WITH INTERACTIONS AMONG SOLUTION ELEMENTS

This third experiment was an effort to replicate and extend the surprising result from the experiment described in Section 3, namely, the group advantage for individual imitation. Members of a group obtained higher scores when their peers had higher, not lower, rates of imitation. A good animal behavior analog for this effect is the cliff swallows studied by Brown and Brown. These birds feed off of airborne insects that travel in large, amorphous clouds that are buffeted by winds. The swallows produce a loud, vocal signal when they have found a region with many insects, even though it is energetically costly to produce this signal, and nearby swallows may compete with the calling bird for insects. One of the reasons why a cliff swallow engages in costly signaling of insect food sources is that it benefits by having other cliff swallows foraging nearby. The swallows recruited by the signal track the subsequent movements of the insects more effectively than the original swallow could if foraging by itself. Likewise, when participants are surrounded by peers who engage in strategic imitation, then the participants may benefit even when their own solutions are being imitated. The imitators will tend to modify what they have imitated, and some of these modifications will produce even better outcomes than the original solution. In these cases, imitated participants can then benefit by reciprocally imitating the peer who originally imitated them.

This experiment (Wisdom, Song, & Goldstone, 2008; *in press*) was designed to offer two methodological improvements over the previous experiment. This previous experiment has the desirable feature that the construction of solutions is relatively open ended, and the final productions look like coherent artistic objects, albeit extremely simple ones. An offsetting disadvantage of these drawings is that it is hard to definitively assess whether a participant is imitating another drawing or simply independently creating similar drawings on their own. Furthermore, the scoring scheme for evaluating the drawings does not permit interactions between solution parts. Interactions between solution parts are particularly interesting because they allow for “rocky” fitness landscapes for a problem space.

Pixel overlap with a hidden “mystery” picture determined fitness, meaning that the quality of a solution was linearly and equally influenced by each pixel. Many real-world problems have a nonlinear characteristic that makes finding good solutions difficult. For evading predators, claws for climbing trees and heavy armor are each good, but the combination is poor because the heavy armor makes climbing difficult. Tea may be improved by either milk or lemon, but not both combined.

The third experiment was designed as a conceptual replication of the second, permitting greater clarity in interpreting participants’ strategies, and greater flexibility and complexity in the design of the search space. We were again interested in documenting the strategies participants used to determine whether to imitate or innovate, and how these strategies affected group-level measures of performance.

4.1. The “Creature League” Task

One hundred and fifty-three participants were distributed across 39 sessions in groups ranging from 1 to 9 participants. Each participant’s task was to score the most points possible over 24 10-s rounds, by assembling together teams of Pokemon-like creatures. [Figure 1.16](#) shows the interface for the experiment. Score feedback was generated according to a stable (within each game) but hidden payoff function, featuring a linear term and pairwise interactions among the icons. In each round, participants could observe each of their fellow players’ solutions and associated scores, and imitate them in whole or in part. The size and the complexity of the problem space (and thus the task difficulty) were manipulated in two different conditions via the sizes of the overall set of icons and the subset that could be evaluated in one solution, as well as the number of pairwise interactions between icons.

All participants’ actions were recorded and synchronized by a game server at the end of each round. The display included an area for the participant’s own current solution (“team” in [Figure 1.16](#)), an area that could be toggled to show the participant’s team on the previous round or their best-scoring team so far in the game (along with its associated score), an area which showed all of the solution elements (the “league” of potential team members) that were available for selection, and indications of the current round in the game and the amount of time remaining in the current round. In sessions with more than one participant, the display also showed the solution and associated payoff of each other participant from the previous round. The ordering of peers’ solutions in each participant’s display was

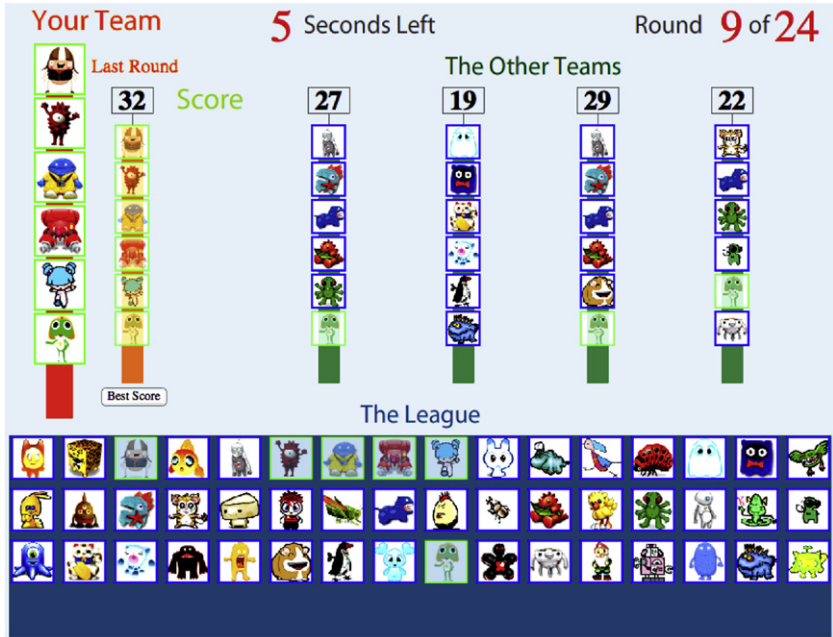


Figure 1.16 Example of experiment interface for experiment 3. Participants assemble their teams on the left. Teams assembled by other peers are shown on the right. The entire league of possible players is shown in the bottom section labeled “the league.” For color version of this figure, the reader is referred to the online version of this book.

kept constant within each condition but not across conditions, to avoid imitation based on past behavior.

Any individual element could be copied from any part of the display to a participant’s current solution by dragging and dropping it with the mouse, except for those already in the participant’s current solution, which were faded in the display and nondraggable. The current solution could be replaced entirely by another solution by selecting the score box above the latter as a “handle” and dragging it to the current solution area. A short video demonstrating all available actions in the game can be viewed at <http://cognitrn.psych.indiana.edu/CreatureGameClip.mov>.

In each game, each creature icon was associated with a certain positive number of points (its own “abilities”), and several unidentified pairs of icons were associated with separate positive point bonuses or negative point penalties (reflecting “how well they got along”) when they were both on the same team in the same round. These latter bonuses

and penalties can be understood as “interaction effects” on top of the “main effects” of each icon’s value, and is how the complexity of the search space was manipulated. Simply adding the influences of each individual icon does not predict a team’s score because of these interaction terms. The icons’ display positions and associations with the payoff function were shuffled randomly for each game, so that their appearance and placement in the display did not give clues as to their point values. Score feedback (the sum of the individual and pairwise terms described above) was given after each round.

4.2. Major Results and Implications

Over the course of the 24 rounds, people’s scores improved substantially, and improved more in the larger group sizes. Imitation of other players’ teams was common, and the score of the imitated participant was greater than that of the imitator in 89.6% of cases, equal to it in 2.6% of cases, and less than that of the imitator in 7.8% of cases, consistent with a “copy better-performing individuals” strategy.

One apparent heuristic that participants use to choose an icon to imitate is to select an icon that is popular among peers. If one were to randomly select an icon to imitate, then an icon that appears on more peers’ teams would be more likely to be selected. The expected probability of an icon being selected by this random imitation strategy is shown by the straight dashed line in Figure 1.17a. In fact, the empirically observed probability of imitation increases more precipitously with an icon’s frequency

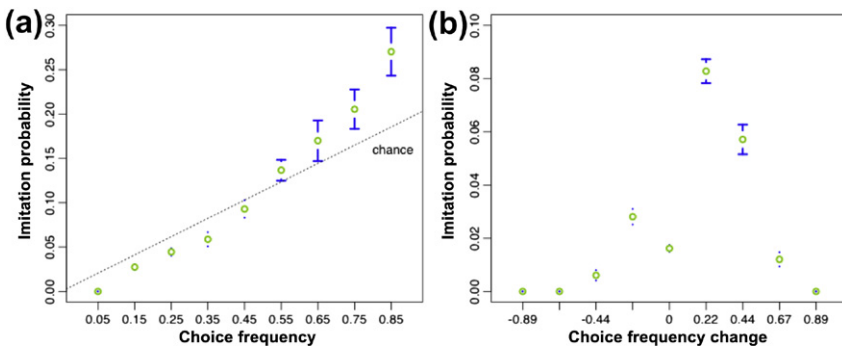


Figure 1.17 (a) Experiment 3 shows a bias toward choosing solution elements that were more frequently represented on other teams. This bias exceeds what would be expected by chance. (b) There was a positive momentum bias toward choosing elements whose representation on other teams was increasing over rounds. For color version of this figure, the reader is referred to the online version of this book.

than the dashed line, indicating that people have an even greater probability of selecting popular icons than would be predicted by chance.

Another kind of heuristic is to use the momentum of an icon's popularity as a cue to the usefulness of an icon. An icon has positive momentum on Round t if it increased in frequency on peers' teams from Round $t-1$ to t . An icon has negative momentum if it decreased in frequency from Round $t-1$ to t . One plausible assumption is that an icon has positive momentum because it has conferred an improvement on these teams. Accordingly, participants may not only be using the base popularity of an icon to select an icon to imitate but also the round-to-round change in popularity, choosing on Round $t+1$ an icon that has positive momentum on Round t . The X-axis of Figure 1.17b is the change in frequency of an icon from one round to the next. The strong deviation from symmetry around the $x=0$ axis indicates a sizeable positive momentum bias. For example, participants are much more likely to choose an icon that has increased its frequency on peers' teams by 0.22 rather than decreased its frequency by 0.22, even equating for the current frequency of the icon.

The second experiment revealed a bias (Figure 1.15) for participants to selectively imitate teams that represented solutions similar to their own solutions. The third experiment replicates this effect, as shown in Figure 1.18. The horizontal axis of Figure 1.18 represents the similarity of two teams of icons. For example, if two teams share 4 out of 6 of their icons, then their similarity is 67%. The vertical axis shows the probability of an event occurring. As the similarity of two teams increases, the probability that the creator of one of these teams will imitate (top panel) rather than ignore (bottom panel) the other team also increases.

Overall, the most common strategy is simply to retain icons on one's team, accounting for 74% of the icons on teams, followed by exploring by selecting icons from the league (15%), followed by imitating icons from other participants' teams (9%). Imitation became increasingly prevalent as the size of a group increased, and decreased with passing rounds. When participants imitated, they were much more likely to copy icons from teams that scored better than their own teams, very often the best available team. Exploring icons from the league also decreased with rounds, and the more conservative strategies of retaining icons and returning to one's own previous teams increased with rounds. Diversity of solutions decreased over rounds, and scores increased. Scores also increased as a function of group size, as larger groups brought in additional innovations for each participant to potentially incorporate in their own solutions.

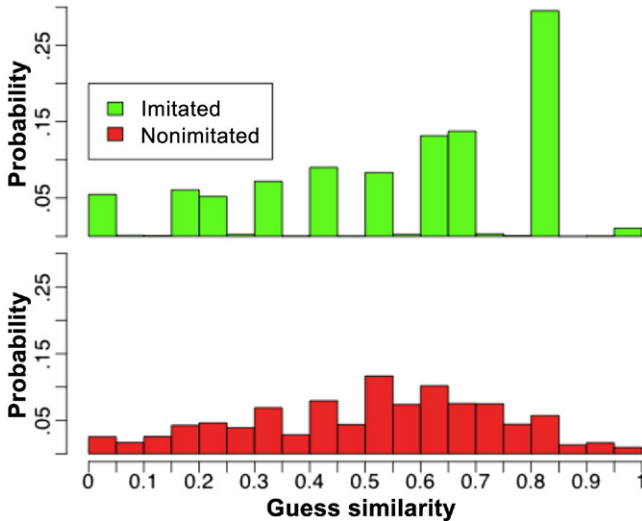


Figure 1.18 Imitators' previous teams showed greater similarity to the teams they imitated than to those they did not imitate. The gaps in histogram distributions are artifacts of the coarse binning of team dyads into similarity values and the fact that not all similarity values are achievable with only six team members. For color version of this figure, the reader is referred to the online version of this book.

From a pragmatic perspective, one might be interested in which strategies lead to the best scores for an individual. Retaining icons previously on one's team produced the best overall score for individuals, followed by imitation, then retrieving previous teams, and lastly exploring. Exploring by sampling unknown icons from the league is a risky strategy, particularly after one has found a team with a score that is substantially better than a random team. However, it is also true that exploratory choices were more prevalent in teams that produced improved scores within the entire group (18% of icons) rather than those that did not offer improvements (13% of icons). A relatively small amount of exploration is collectively useful for bringing in new possibilities. As individuals imitated more, regressions indicated that their scores were likely to increase, and a similar positive relation was found for the retention strategy.

Very similar and significant patterns of results were shown in analyses of mean group score versus mean group guess proportion for each choice source, even when each individual was excluded from their group's aggregate behavior. That is, an *individual's score* was higher when *the individual's fellow group members* imitated and retained more, and explored and retrieved less. This pattern replicates the second experiment's surprising pattern that

it is better for an individual to be surrounded by imitators. A complete lack of exploration will of course result in a lack of improvements, but this experiment suggests that in a large and complex problem space, productive exploration may be readily incentivized by the potential for generating small improvements based on peers' solutions. This is analogous to the mixed equilibrium for individual contributions to group efforts found by [Kameda and Tindale \(2006\)](#).

The results regarding the group benefits of imitation and collective risks of exploration, taken together with the reductions in diversity over rounds, imply a view that is at odds with those predicted from a simple producer-scrounger dilemma interpretation of social learning ([Kameda & Nakanishi, 2003](#)). Much like "conformity," being a "scrounger" often carries a negative connotation or denotation, such as "social loafing" ([Latané, Williams, & Harkins, 1979](#)). However, such behavior may be appropriate when not all group members' full efforts are required to produce sufficient benefit. In a complex but relatively stable environment, the best outcome for the group may result from most group members converging on a "good enough" solution quickly to achieve high mean performance, and then introducing productive exploration when necessary. Given a baseline inclination to some amount of individual exploration, the limiting factor in improving search performance may be the amount of information sharing and coordination among searchers, which allow them to pool both the benefits and the risks of asocial learning ([Hess & Ostrom, 2007](#)).



5. LIMITATIONS, IMPLICATIONS, AND CONCLUSIONS

There are certainly limitations to the external validity of the reported experiments. The kinds of "innovations" that our participants were engaged in discovering were minor, simple, and highly constrained compared to the innovations created by artists, pharmaceutical companies, and even elementary school students during recess. Our participants only worked on revising their solutions for at most an hour, and each solution could be expressed in only 10–20 s. Given these limitations, it would be foolhardy to draw major implications from our studies for cultural improvement at a societal scale.

Perhaps the strongest general conclusion that can be drawn from our work is simply that social learning is a major factor in people's performance and problem-solving capacity. In this respect, the limited nature of our experimental paradigms is a rhetorical strength. Even when broader cultural contexts are kept to a minimum with our laboratory-based experiments,

participants often imitate one another and do so in predictable ways. Furthermore, even when the groups to which participants belong are ad hoc and temporary, and have only rudimentary communication possibilities, there are still sufficiently rich interactions between group members for unexpected group-level phenomena to arise. If these phenomena are robust enough to be found in our constrained, well-controlled laboratory conditions, then there is good reason to expect them to occur in other real-world contexts as well.

5.1. Imitation Heuristics

Some of the heuristics for imitation that we observed have been previously documented in animal behavior and social psychology. Others are more novel. Across the three reported experiments, the heuristics for which we have solid, replicated evidence include the following:

5.1.1. Frequency

Imitate options that are relatively prevalent among one's peers. In our experiments, this frequency heuristic led to choices of frequently occurring options more than would be predicted by chance. For example, in the last experiment, icons that were the most prevalent, representing 17% of all icons across all teams, were copied by participants who did not already possess the icons 27% of the time.

5.1.2. Upgrade

Imitate options that produce results better than one's existing solution. In our experiments, this takes the form of imitators choosing options that offer higher scoring solutions than the imitator's previous solution. Very often, the imitators choose the highest scoring option available to them.

5.1.3. Early Imitation

Imitate others' options more during the early, compared to late, rounds of innovation search. Early imitation is advantageous because one's own solutions are less likely to be strong at first, and there will be considerable diversity among solutions. In addition, uncertainty about a problem space is the largest at the beginning of the search process (Kendal, Cooley, & Laland, 2009). With passing rounds, our participants became more committed to their own solutions and were less likely to take a radically different approach by imitating a dissimilar solution. In our experiments, both imitation and open-ended exploration decreased over rounds, and although

these strategies have been contrasted, both are strategies for increasing the diversity of one's own solutions. As more information is gained about what solutions work well in a domain, the more conservative strategies of retaining one's solution and returning to one's previously strong solutions become more prevalent.

5.1.4. Similarity

Imitate elements of solutions that are already similar to one's own solution. All three experiments provide evidence that participants tended to preferentially copy relatively similar solutions. One reason why this similarity heuristic may be adaptive is that it prevents incorporating solution elements that are incompatible with one's previous solution and knowledge of the problem space. A bias toward borrowing from similar rather than dissimilar solutions has also been incorporated into general machine learning algorithms featuring multiple agents simultaneously searching for solutions (Goldberg, 1989, chap 1, pp. 1–23; Goldstone & Janssen, 2005). There are two possible drawbacks when agents borrow solution elements from other agents pursuing substantially different solutions. First, they abandon the knowledge of the problem space accumulated in their previous solution. Second, there is a strong risk that the resulting blend of solutions will be a suboptimal hybrid not well adapted to the niche of either of the original solutions. Given the complex search landscapes used in the experiments, participants may have been biased to copy solution elements from similar rather than dissimilar solutions to ensure greater solution compatibility.

5.1.5. Momentum

Imitate options that are increasing in their prevalence. The last two experiments revealed that participants tended to select options not simply based on their frequencies in the population, but also based on their round-to-round change in frequency. Options that have an increasing "market share" tend to be selected, and options that have a decreasing market share tend to be avoided. Our participants may have been using positive momentum as cue to the beneficial consequences of having a particular solution element.

Of these heuristics, the last two are the most novel, but even these have some precedent. In a form of Similarity Heuristic, Rogers (2003) observed that people, companies, and institutions often adopt innovations that are compatible with the solutions that they already employ. Our experimental contribution here is to show that this Similarity Heuristic continues to be adopted even when there are no retooling costs. Our experimental

evidence for a Similarity Heuristic despite all solutions being equally easily adoptable suggests that at least part of the basis for this heuristic is cognitive inertia and the tendency for people to preferentially continue pursuing their own approaches.

A real-world precedent for the Momentum Heuristic is shown in baby-naming decisions by parents. An examination of 130 years of social security data on baby names reveals that, for the last 60 years, baby names that increase in popularity from 1 year to the next tend to increase in popularity still further in the following year (Gureckis & Goldstone, 2009). Likewise, a decrease in popularity tends to be followed by still further decreases. This stands in contrast to the years 1880–1940, when increases in popularity were more likely to be followed by decreases, and decreases by increases. The United States has gradually switched from a negative to positive momentum society, at least in terms of its baby names. It is tempting to suggest that this reflects an increasing “faddishness” in American society. Parents, wishing to avoid giving their child a name that will be unpopular in the future, use the increasing momentum of a name as a cue to its future popularity. An unintended consequence of employing a Momentum Heuristic is that the distribution of options in a group becomes increasingly well predicted by a Momentum Heuristic. As our participants or American parents increasingly rely on the Momentum Heuristic to make their choices, then the distribution of choices in the group becomes increasingly well predicted by positive momentum, further justifying the use of the Momentum Heuristic if one’s aim is to predict future popularity.

5.2. Group-level Phenomena

This unintended consequence of the Momentum Heuristic is a striking example of a group-level phenomenon that emerges from individual social learning heuristics. It is characteristic of the kind of collective phenomena that arise when decision makers affect the environment for subsequent decision makers because the environment is largely comprised of their decisions (Goldstone & Gureckis, 2009; Goldstone & Roberts, 2006; Roberts & Goldstone, 2006). One of our primary interests has been in bridging between individual- and group-level phenomena. Some of the major group-level phenomena that we have observed are as follows:

5.2.1. Convergence

Members of a group tend to converge on similar solution with time if they see each others’ solutions and measures of the success of these solutions.

5.2.2. Inefficient Problem Space Coverage

As a direct result of convergence, a group will often not efficiently cover a problem space. This was most striking in the first experiment, in which groups with members who could see every other member's solution performed less well than groups with members who had restricted access to others' solutions.

5.2.3. Problem Space/Social Knowledge Match

How well a group as a whole will solve a problem will depend on the match between the complexity of the problem space and its members' access to others' solutions. The specific nature of this interaction was revealed by human experiments, and broadly corroborated by computer simulations. In general, as the global maximum of a problem space becomes increasingly difficult to find via a simple hill-climbing search, then increasingly restricting the visibility of peer solutions will promote group performance. For easy problems, the critical determinant of group performance is how quickly word can be spread about good solutions, and hence, broadly interconnected social networks are superior. For hard problems, more sparsely interconnected social networks help the group explore different solution possibilities in parallel.

5.2.4. Reciprocal Imitation

Members of a group can benefit by being imitated because the imitators will subsequently modify the solution, and if the modification is favorable, the imitated members can copy the improved solution. One of the most surprising results from the last two experiments was that individuals benefited from being in groups with others who frequently imitated. Reciprocal imitation is a large part of the reason for this benefit. Imitation also has the collective benefit of keeping good solutions alive in the collective memory (Wegner, Erber, & Raymond, 1991).

Future work will be necessary to determine when different group-level patterns are observed. There is some tension between the benefits of having sparsely connected groups in the first experiment and the benefits of being surrounded by imitators observed in the last two experiments. Our tentative reconciliation is that being imitated is particularly advantageous when one is trying to search a region of a problem space that is too large for one to effectively survey by oneself. For the huge search spaces of the later experiments, it is to one's advantage to recruit others to one's region because these recruits can assist in the regional search for better solutions. In

the first experiment, one's chances for finding the global maxima are better if other people search in different areas rather than redundantly searching in the same region of the one-dimensional problem space.

While the external validity of our current experiments is undeniably limited, we take it as a sign that we are on a promising track that issues of external validity even arise. Humans are almost always social learners. We learn by being told, by being shown, and by watching others who are simply behaving and not trying to demonstrate anything. At a societal level, this social learning produces important and striking group-level consequences, including bandwagons, speculative bubbles, schisms and coalitions within a group, spontaneous formation of minority-opinion groups, opinion cycles, group polarization of opinion, and early market advantages. The kinds of solution copying processes that we observe in our experiments can hopefully provide a set of core patterns that combine in different ways to create these large-scale social patterns that intrinsically matter to us and shape our experiences and well-being. For this reason, the spread of solutions in a group is a process of consequence for the psychology of both learning and motivation.

ACKNOWLEDGMENT

The authors wish to express thanks to Katy Borner, Todd Gureckis, Brian Ross, Rich Shiffrin, Eliot Smith, William Timberlake, and Peter Todd, for helpful suggestions on this work. This research was funded by the Department of Education, Institute of Education Sciences grant R305H050116, and National Science Foundation grant 0527920. More information about the laboratory and online versions of experiments related to the current work can be found at <http://cognitnrn.psych.indiana.edu>.

REFERENCES

- Acheson, J. M. (2003). *Capturing the commons: Devising institutions to manage the Maine lobster industry*. Lebanon: University Press of New England.
- Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small groups as complex systems: Formation, coordination, development, and adaptation*. Newbury Park, CA: Sage.
- Asch, S. E. (1956). Studies of independence and conformity: a minority of one against a unanimous majority. *Psychological Monographs*, 9, (Whole No. 416).
- Atran, S., Medin, D. L., & Ross, N. (2004). Evolution and devolution of knowledge: a tale of two biologies. *Journal of the Royal Anthropological Institute*, 10, 395–420.
- Axelrod, R. (1997). The dissemination of culture: a model with local convergence and global polarization. *Journal of Conflict Resolution*, 4, 203–226.
- Bettencourt, L. M. A. (2009). The rules of information aggregation and emergence of collective intelligent behavior. *Topics in Cognitive Science*, 1, 598–620.
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100, 992–1026.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.
- Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. New York: Oxford University Press.

- Bullnheimer, B., Dawid, H., & Zeller, R. (1998). Learning from own and foreign experience: technological adaptation by imitating firms. *Computational and Mathematical Organizational Theory*, 4, 267–282.
- Cacioppo, J. T., & Patrick, W. (2008). *Loneliness: Human nature and the need for social connection*. New York: W. W. H. Norton & Co.
- Chwe, M. S.-Y. (1999). Structure and strategy in collective action. *American Journal of Sociology*, 105, 128–156.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annual Review of Psychology*, 55, 591–621.
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informative social influences upon individual judgment. *Journal of Abnormal Social Psychology*, 51, 629–636.
- Dodds, P., Muhamad, R., & Watts, D. (2003). An experimental study of search in global social networks. *Science*, 301, 827–829.
- Dorogovtsev, S. N., & Mendes, J. F. F. (2003). *Evolution of networks: From biological nets to the internet and WWW*. Oxford: Oxford University Press.
- Feynman, R. P., Leighton, R., & Hutchings, E. (1997). “Surely you’re joking, Mr. Feynman!”: *Adventures of a curious character*. New York: W. W. Norton & Company.
- Galef, B. G., & Laland, K. N. (2005). Social learning in animals: empirical studies and theoretical models. *BioScience*, 55, 489.
- Gigone, D., & Hastie, R. (1996). The impact of information on group judgment: a model and computer simulation. In J. H. Davis, & E. Witte (Eds.), *Understanding group behavior: Consensual action by small groups* (Vol. 1, pp. 221–251). Hillsdale, NJ: Erlbaum.
- Gladwell, M. (2000). *The tipping point*. London: Little Brown and Company.
- Goldberg, D. E. (1989). *Genetic algorithms*. Reading, MA: Addison-Wesley.
- Goldstone, R. L., & Ashpole, B. C. (2004). Human foraging behavior in a virtual environment. *Psychonomic Bulletin & Review*, 11, 508–514.
- Goldstone, R. L., & Gureckis, T. M. (2009). Collective behavior. *Topics in Cognitive Science*, 1, 412–438.
- Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. *Trends in Cognitive Science*, 9, 424–430.
- Goldstone, R. L., & Roberts, M. E. (2006). Self-organized trail systems in groups of humans. *Complexity*, 15, 43–50.
- Gureckis, R. L., & Goldstone, R. L. (2009). How you named your child: understanding the relationship between individual decision making and collective outcome. *Topics in Cognitive Science*, 1, 651–674.
- Henion, M., & Fischhoff, B. (1986). Assessing uncertainty in physical constants. *American Journal of Physics*, 54, 791–797.
- Hess, C., & Ostrom, E. (Eds.), (2007). *Understanding knowledge as a commons*. Cambridge, MA: MIT Press.
- Hills, T. T., Todd, P. M., & Goldstone, R. L. (2010). The central executive as a search process: priming exploration and exploitation across domains. *Journal of Experimental Psychology: General*, 139, 560–609.
- Hurley, S., & Chater, N. (2005). *Perspectives on imitation: From mirror neurons to memes*. Cambridge, MA: MIT Press.
- Kahan, J. P., & Rapoport, A. (1984). *Theories of coalition formation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kameda, T., & Nakanishi, D. (2003). Does social/cultural learning increase human adaptability? Rogers’ question revisited. *Evolution and Human Behavior*, 24(4), 242–260.
- Kameda, T., & Tindale, R. S. (2006). Groups as adaptive devices: human docility and group aggregation mechanisms in evolutionary context. In M. Schaller, J. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 317–342). New York: Psychology Press.
- Kearns, M., Suri, S., & Montfort, N. (2006). An experimental study of the coloring problem on human subject networks. *Science*, 313, 824–827.

- Kendal, R. L., Coolen, I., & Laland, K. N. (2009). Adaptive trade-offs in the use of social and personal information. In R. Dukas, & J. M. Ratcliffe (Eds.), *Cognitive ecology II* (pp. 249–271).
- Kennedy, J. (2009). Social optimization in the presence of cognitive local optima: effects of social network topology. *Topics in Cognitive Science*, 1, 498–522.
- Kennedy, J., Eberhart, R. C., & Shi, Y. (2001). *Swarm intelligence*. San Francisco: Morgan Kaufmann Publishers.
- Kerr, N., & Tindale, R. (2004). Group performance and decision making. *Annual Review of Psychology*, 55, 623–655.
- Kleinberg, J. (2000). Navigation in a small world. *Nature*, 406, 845.
- Laland, K. N. (2004). Social learning strategies. *Learning & Behavior*, 32, 4–14.
- Latané, B., & Bourgeois, M. J. (1996). Dynamic social impact in electronic groups: consolidation and clustering in computer-mediated communication. *Journal of Communication*, 46, 35–47.
- Latané, B., & L'Herrou, T. L. (1996). Spatial clustering in the conformity game: dynamic social impact in electronic groups. *Journal of Personality and Social Psychology*, 70, 1218–1230.
- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: the causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37, 822–832.
- Lazer, D., & Friedman, A. (2005). *The parable of the hare and the tortoise: Small worlds, diversity, and system performance*. KSG Working Paper No. RWP05–058. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=832627.
- Lerner, J., Tirole, J., & Pathak, P. (2006). The dynamics of open source contributors. *American Economic Review Papers and Proceedings*, 96, 114–118.
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of the crowd effect. *Proceedings of the National Academy of Sciences*, 108, 9020–9025.
- Mason, W. A., Jones, A., & Goldstone, R. L. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology: General*, 137, 422–433.
- Mason, W. A., & Watts, D. J. (2011). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, 109, 764–769.
- Milgram, S. (1967). The small world problem. *Psychology Today*, 2, 60–67.
- Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper Collins.
- Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: a dynamic theory of social impact. *Psychological Review*, 97, 362–376.
- Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, games, and common-pool resources*. Ann Arbor: University of Michigan Press.
- Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., et al. (2010). Why copy others? Insights from the social learning strategies tournament. *Science*, 328, 208–213.
- Roberts, M. E., & Goldstone, R. L. (2006). Epicure: spatial and knowledge limitations in group foraging. *Adaptive Behavior*, 14, 291–313.
- Roberts, M. E., & Goldstone, R. L. (2011). Adaptive group coordination and role differentiation. *PLoS One*, 6, 1–8.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Rosenkopf, L., & Abrahamson, E. (1999). Modeling reputational and information influences in threshold models of bandwagon innovation diffusion. *Journal of Computational and Mathematical Organizational Theory*, 5, 361–384.
- Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311, 854–856.
- Salganik, M. J., & Watts, D. J. (2009). Web-based experiments for the study of collective social dynamics in cultural markets. *Topics in Cognitive Science*, 1, 439–468.

- Sang, K., Todd, P. M., & Goldstone, R. L. (2011). Learning near-optimal search in a minimal explore/exploit task. In *Proceedings of the thirty-third annual conference of the cognitive science society* (pp. 2800–2805). Boston, Massachusetts: Cognitive Science Society.
- Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143–186.
- Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, *187*, 60.
- Simon, H. A. (1957). *Models of man, social and rational*. New York: Wiley.
- Strang, D., & Macy, M. (2001). In search of excellence: fads, success stories, and adaptive emulation. *American Journal of Sociology*, *107*, 147–182.
- Sumpter, D. J. T. (2010). *Collective animal behavior*. Princeton, NJ: Princeton University Press.
- Templeton, J. J., & Giraldeau, L. A. (1996). Vicarious sampling: the use of personal and public information by starlings foraging in a simple patchy environment. *Behavioral Ecology and Sociobiology*, *38*, 105–114.
- Theiner, G., Allen, C., & Goldstone, R. L. (2010). Recognizing group cognition. *Cognitive Systems Research*, *11*, 378–395.
- Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*, 495–552.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*, 440–442.
- Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, *61*, 923–929.
- Wisdom, T. N., & Goldstone, R. L. (2011). Innovation, imitation, and problem-solving in a networked group. *Nonlinear Dynamics, Psychology, and Life Sciences*, *15*, 229–252.
- Wisdom, T. N., Song, X., & Goldstone, R. L. (2008). The effects of peer information on problem-solving in a networked group. In *Proceedings of the thirtieth annual conference of the cognitive science society* (pp. 583–588). Washington, D.C.: Cognitive Science Society.
- Wisdom, T. N., Song, X., & Goldstone, R. L. (in press). Social Learning Strategies in a Networked Group. *Cognitive Science*.
- Wolff, P., Medin, D. L., & Pankratz, C. (1999). Evolution and devolution of folk biological knowledge. *Cognition*, *73*, 177–204.

This page intentionally left blank



Space, Time, and Story

Barbara Tversky*,¹ Julie Heiser[†], Julie Morrison[‡]

*Teachers College, Columbia University, New York, NY, USA; Department of Psychology, Stanford University, Palo Alto, CA, USA

[†]Adobe Systems, San Jose, CA

[‡]Department of Psychology, Glendale Community College, Glendale AZ

¹Corresponding author: E-mail: btversky@stanford.edu

Contents

1. Introduction: Life, Perception, and Stories	47
2. Perceiving the Events of Life	48
3. Telling the Events of Life: Word, Diagram, Gesture	53
3.1. Routes: Nodes and Links	53
3.2. Assembly: Step-by-Step, Perspective, Action	54
3.3. Visual Story Semantics: Depiction, Description, and Gesture	60
3.4. Principles of Apprehension and Congruence: Animation?	61
3.5. Static Tales: Word and Picture	63
3.6. Making Inferences	64
3.6.1. <i>Designers Make Discoveries in Their Own Sketches</i>	64
3.6.2. <i>Making Inferences from Diagrams and Descriptions</i>	65
3.7. Gesture for Action	68
3.8. 'Real' Stories	70
3.9. Explanations and Narratives	71
4. Spraction: Space, Action, Abstraction	72

Abstract

Life presents as a continuous multimodal barrage on all our senses. From that, we abstract events, discrete units characterized by completion of goals and peaks of action. Effective communication of sequences of events in explanations and narratives is similarly segmented, and linked globally by overall themes and locally by anaphora. Visuospatial explanations and narratives, notably diagrams, comics, and gestures, rely on congruity of mappings of elements and relations of ideas to space and marks in space. Just as we design visuospatial discourse, we design the world: Our design actions in space create diagrams in the world, patterns, piles, rows, one-to-one correspondences, and the like, that express abstractions, categories, hierarchies, dimensions, and more, a circular process termed *spraction*.



1. INTRODUCTION: LIFE, PERCEPTION, AND STORIES

There is life. There is the perception of life. There are stories of life.

Life just keeps happening: unbroken, continually, continuously, ubiquitously, and inexorably. Life happens in sight, in sound, in smell, and in touch:

all at once and from all directions. Life is outside. It happens in space and time. It happens without narration and without explanation.

Perception of life is quite different. Perception is inside. Perception goes beyond the information given. Perception happens in lumps and pieces and pieces of pieces. The pieces are multiply organized across modalities, in space and in time. Perception of life happens with interpretations. Interpretations link the pieces.

Stories of life are different yet again. Stories are outside again. Stories are again multiply organized, in time, in space, and more. Stories have a point of view, an audience, an agenda. Stories, especially in the wild, can be told with words, with gesture, with prosody, with props, and with sketches.

What follows is a story of stories: a story about the perception and cognition of the events of life, a story about stories of life, a story about story telling. Finally, a story called *Spraction* that integrates space, action, and abstraction; diagrams, gesture, and thought. Because this is a story (of stories), it has a point of view, an agenda, and an audience. Its point of view is mine, its agenda is surveying some of the research my stimulating collaborators and I have been involved in, and its audience is you. I hope you will find the work almost as exciting and insightful as we have.

Truth in advertising. The stories will not be the gripping stories in books or films, or those told by your friends. The stories will be simple and mundane; they will take place in space and be ordered in time. They usually have action and causes. Most of the stories will be visual and most will be explanations; nevertheless, they carry many of the key features of traditional stories. They have patterns, though it is often the breaking or expanding of those basic patterns that allow the drama and surprises and suspense of gripping stories. For the most part, the stories told here lack characters and lack social intrigue. Despite those caveats for modesty, we begin on a grandiose scale: perceiving the events of life.



2. PERCEIVING THE EVENTS OF LIFE

How is life perceived? Jeff Zacks, Bridgette Martin Hard, a few others, and I have been studying the perception of life, or snippets of life (Hard, Recchia, & Tversky, 2011; Tversky & Zacks, *in press*; Tversky, Zacks, & Hard, 2008; Tversky, Zacks, Morrison, & Hard, 2010; Zacks and Tversky, 2001; Zacks, Tversky, & Iyer, 2001). We chose snippets that are common and familiar, that do not take much time, that could be filmed from a single vantage point: snippets like making a bed or assembling a piece of furniture, or getting ready for work in the morning. For some studies, we chose snippets

that were abstract and unfamiliar, such as geometric figures moving around enacting scenarios like taunt bully and chase and hide and seek. Clearly, the kinds of snippets we studied do not have the full complexity and chaos of life, but they are representative of large swatches of life.

Our initial studies on familiar events (Zacks et al., 2001) and some of the subsequent ones adapted a paradigm of Newtson (Newtson, 1973; Newtson & Engquist, 1976; Newtson, Engquist, & Bois, 1977). Observers watched the videos, and pressed a button when they thought one event segment ended and another began. Typically, they did this twice, once at the coarsest level that made sense and once at the finest level that made sense. Some were asked to give play-by-play descriptions; to say at every button press, what had happened in that segment. Others were asked to give the play by play after they had viewed the videos.

If life were perceived as an unbroken continuous continual multisensory change, then people would press buttons at random and idiosyncratically. They did not. On the whole, people's button presses fell at approximately the same places. What is more, button presses were organized hierarchically, that is, the coarse presses—the standard term is 'breakpoint'—coincided with the fine breakpoints far greater than chance. Thus, breakpoints were organized temporally, and also spatially, at the same point in time. More on that later. Now, another coincidence is that according to the play-by-play descriptions, the breakpoints were at places where goals were completed, coarse breakpoints for larger goals, and fine breakpoints for subgoals. Here is part of one transcript for coarse breakpoints for *making the bed*: walking in; taking apart the bed; putting on the sheet; putting on the other sheet; putting on the blanket. The same participant, describing what happened at the fine breakpoints for the coarse breakpoint, *putting on the sheet*: unfolding the sheet; laying it down; putting on the top end of the sheet; putting on the bottom; straightening it out. These transcripts illustrate another convergence: new objects (or object parts) segmented events at the coarse level, and new actions on the same object (or object part) segmented events at the fine level. Intriguingly, in other studies where participants segmented at only one level, that level corresponded best to the coarse level, the level where new objects coincide with new segments. How general the later phenomena are remains to be seen.

However, the play-by-play descriptions yielded another general and revealing phenomenon about how events are perceived and comprehended. Almost without exception, the descriptions included a beginning, a play-by-play middle, and an end. This is of course a fundamental characteristic of stories, they have beginnings, middles, and ends; they start somewhere, something happens, they end somewhere else.

Up till now, we have seen that observers reliably segment events hierarchically at points where goals and subgoals are completed. This might suggest that the top-down understanding of events drives segmentation. That this is conjecture is far from complete has been shown by subsequent work. As for so many phenomena, it goes both ways, both top down and bottom up, the cognition and the perception, converge on event boundaries and provide entries to them. Breakpoints turn out to be the places where physical change is locally maximal, implicating a role for bottom-up perceptual information in event segmentation. Recall that we also studied abstract events that were not easy to interpret, geometric figures enacting bully and chase and hide and seek (Hard, Tversky, & Lang, 2006). As before, some participants were asked to segment and interpret and others were asked to interpret. The amount of movement in each 1-s interval was also computed. The movement score correlated well with breakpoints, with greater amounts of movement at coarse breakpoints, lesser at fine. However, the interpretations indicated that participants understood the events only at the level of movement; they described the actions using terms like stop, start, and rotate. In a second condition, observers viewed the videos five times before segmenting, instead of just a single time. After they had seen the videos five times, they discerned the structure of the events, and described the actions using intentional words, like bully, taunt, hide, and search. Nevertheless, the segment boundaries were the same, at points where movement was high, for both groups, those who segmenting after viewing the videos only once and had low-level interpretations of the motion, and those who segmented after viewing the videos five times and interpreted the actions in terms of goals and intentions.

The confluence of top-down conceptual information and bottom-up perceptual information was corroborated in a subsequent set of studies, using naturalistic events and another index of motion (Hard et al., 2011). These studies also demonstrate another important property of breakpoints, of segment boundaries: They attract extra attention. This makes sense if more is happening, if there is more change at breakpoints, and that turns out to be the case. If more is happening, more attention is needed to comprehend what is happening. Those are some of the conclusions. Now to the bases for these conclusions. In the first of the set of studies, participants watched slide shows sampled every second from videos of naturalistic events, such as having breakfast or cleaning up a messy bedroom. They were told their memory would be tested and that they were free to view each slide for as long as they liked. After the memory test, they were asked to

segment the videos from which the slide shows were taken. This allowed direct comparison of their breakpoints to their looking times. In a separate computational study, the pixel-to-pixel change between each slide and the slide that came before and after was determined, on images that had been filtered to enhance edges and consequently object outlines. Importantly, the three measures converged. Looking times were maximal at breakpoints, longer for coarser breakpoints and shorter for finer ones. Looking times were maximal at moments of greatest change, and the change index was higher for coarser breakpoints than for finer ones. The data from [Hard, et al. \(2011\)](#) are reproduced [Figure 2.1](#). In this experiment, participants were free to segment at as many levels as they liked, though most chose to segment at only two levels.

Why are breakpoints, the points at which one segment ends and another begins, points of locally maximal change? These are points of transition

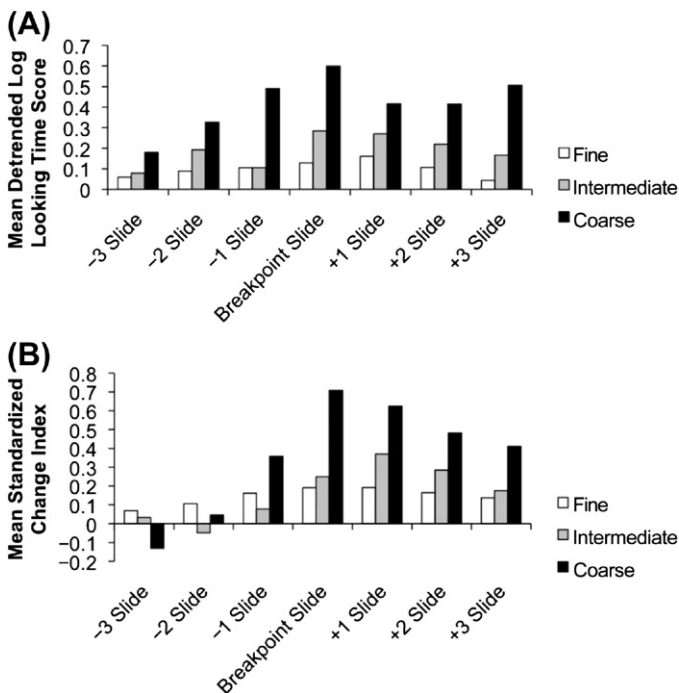


Figure 2.1 A) Mean detrended log looking time scores for participant-defined breakpoint slides at three levels of grain and for immediately preceding and subsequent slides. (B) Standardized pixel-to-pixel change scores for participant-defined breakpoint slides at three levels of grain and for immediately preceding and subsequent slides. (Taken from [Hard, Recchia, and Tversky \(2011\)](#) with permission).

from completing one goal to initiating another, that is, both endings and beginnings. As both endings and beginnings, they incorporate aspects of both, for example, laying down one object and picking up another. That duality, ending and beginning, is bound to entail relatively larger changes in the position of the body, the head, and the limbs, and consequently draws more attention.

Thus, there are good reasons for the convergence of top-down and bottom-up information at breakpoints. Completing one goal and beginning another entails greater movement, hence greater moment-to-moment change in the ongoing physical information. This means that bottom-up and top-down information can predict one another. Knowing that a goal has been completed predicts greater change; greater change signals completion of a goal. Importantly, greater relative change and consequently greater attention organizes the physical input in ways that promote both understanding of ongoing events and acquisition of new events.

Although there was perfect convergence of the three measures, breakpoints, local maxima of change, and looking time, there was one intriguing divergence prior to breakpoints. Looking time began to increase a few seconds before physical change increased. We can only speculate as to why, but other research has shown that actors turn attention and their heads to the next step in an action sequence as they are finishing the current step. For example, in making a sandwich, actors turn their heads to look for the second slice of bread as they finish spreading the first slice (Mennie, Hayhoe, & Sullivan, 2007). Because those shifts of head and eye signal shifts of intention, they are important for predicting what the actor will do next and important for understanding the set of actions. Shifts of the head and eyes are likely to involve only small movements, smaller than the movements in the switch to a new action unit, but likely to attract looking time and attention because they predict the next action unit.

The remarkable confluence of top-down and bottom-up information in segmenting and comprehending events, of breakpoints in action, local maxima of physical change, and local maxima of attention, has several implications. Sharp changes in the physical stimulus can be used to control attention and to organize and acquire unfamiliar events. Completion of goals and subgoals can also serve to control attention and to predict what will happen in the near future. Because breakpoints are determined both perceptually and conceptually, they are natural boundaries between event segments. As natural boundaries, breakpoints are a good place to segment retellings of events, explanations, and stories, whatever the modality of retelling. This is

an appropriate breakpoint to end the discussion of perception of the events of life and begin a discussion of retelling the events of life. Remember also that events are perceived to have beginnings and endings as well as segmented middles.



3. TELLING THE EVENTS OF LIFE: WORD, DIAGRAM, GESTURE

3.1. Routes: Nodes and Links

We begin this part of the journey with one of the simplest stories that people tell from their lives, how to get from here to there. Along with others, notably Michel Denis and his collaborators, we caught people in the streets and asked that question, where “there” was a place known to them (Denis, 1997; Denis, Pazzaglia, Cornoldi, & Bertolo, 1999; Tversky & Lee, 1998, 1999). In our case, we asked some to sketch maps and others to write down directions (Tversky & Lee, 1998, 1999). Notably, both ways of telling routes, depictions and descriptions, had the same underlying structure, suggesting that they derived from the same underlying cognitive structure. Notably again, that structure had a beginning and an end, with a middle that was segmented. The segments corresponded to actions, typically turns, at landmarks, typically intersections. The key to the directions then was turns and landmarks. The exact distances were not crucial as the next landmark would signal when to take a new action. The exact direction of the turn was not crucial as the location of the next link would determine the degree of turn. In fact, people’s sketch maps exaggerated short distances with many actions and minimized distances with no action. The maps distorted directions as well as distances, so that most turns were represented as more or less right angles. Interestingly, these same distortions are present in peoples’ memories of environments, rectifying angles, exaggerating distances that have many events (e.g. Tversky, 1981, 2000a, 2003).

Visually, abstractly, the directions form networks, nodes for landmarks, links for the paths between them. Networks are undoubtedly the most versatile visualization of information. Nodes can stand for any idea and links for any connection. Semantic networks link concepts, social networks link people, computer networks link computers, knowledge networks link ideas. Family trees, decision trees, corporate hierarchies, taxonomic hierarchies are networks, too, with extra constraints, notably asymmetric links, that are labeled, implicitly or explicitly: temporal for families and decisions, power for corporations, is-a for taxonomies. The appeal of networks

as visualizations is dual: They are firmly grounded in space and they can represent the abstract. In using networks, spatial reasoning can be applied to abstract reasoning. As we will see, this duality will hold for other fundamental visualizations of thought.

Proximity still matters (in case anyone thought otherwise) even in the age of ubiquitous connectivity. A pair of computer scientists down the street read the work on spatial cognition and came up with a great idea: to provide intuitive route maps over the web (Agrawala & Stolte, 2001). Until then, map websites gave users maps annotated with the routes requested. Such maps are not very useful because they are at a single scale; hence, critical information such as getting on and off of highways is hard to see and trivial information such as long distances on highways is easy to see. Agrawala and Stolte extracted the cognitive design principles implicit in people's sketch maps and memories: Exact distance does not matter; exact direction does not matter; what matters is where to turn and how. They applied those principles (and more) to develop an algorithm to produce route maps on demand. These were a huge success, and inspired us to find a new project that used cognitive research to reveal cognitive design principles and instantiated those principles in computer algorithms to produce diagrams on demand.

3.2. Assembly: Step-by-Step, Perspective, Action

For the new project, we chose another oft-told tale in our lives, instructions to put something together, from origami and Lego to sound systems or a barbecue (Figure 2.2). Everyone has tales of woe of incomprehensible instructions or extra parts. Instructions to operate something, from a toy airplane to a sophisticated camera, explanations of how something works, from hearts to engines, are similar in structure, and in problems. We thought that in many cases, users, newly experienced ones, would do better than whoever writes the instructions. Users as designers, a program we later adopted more widely (e.g. Kessell & Tversky, 2011). Julie Heiser combed the local big box stores and returned with boxes of unassembled TV carts (Figure 2.3).

These were easy enough for most of our participants, Stanford students, to assemble in a single experimental session, even without instruction, using the photograph on the box. In a series of experiments, we first asked students to assemble the TV cart to acquire the requisite expertise. They then produced instructions to help others with the same task, using, in different experiments, only diagrams or diagrams and language or only language or only gestures. We also measured spatial ability, using the Vandenburg and Kuse

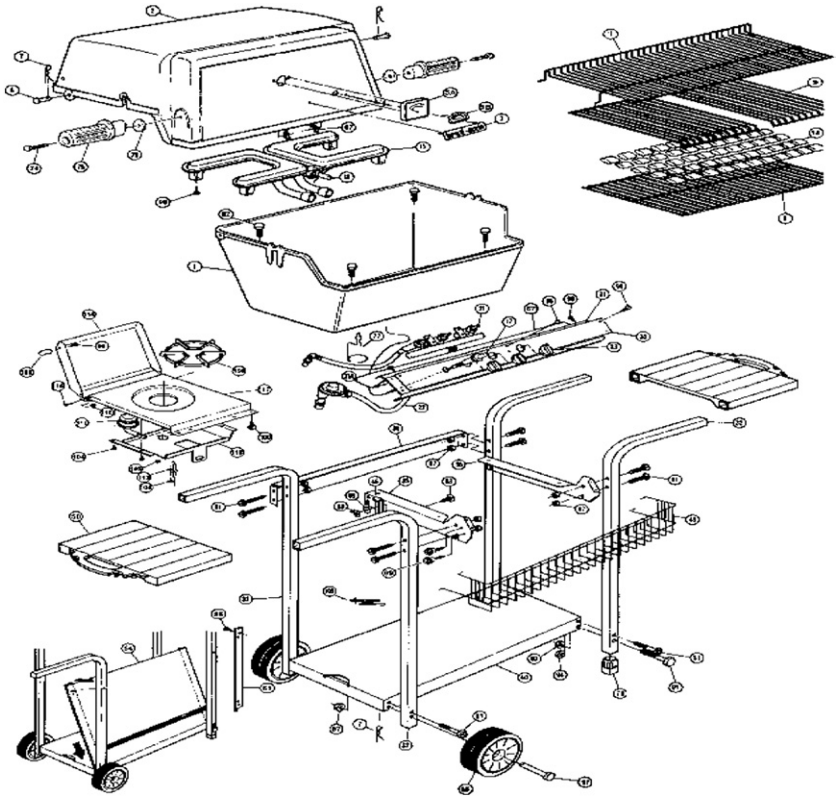


Figure 2.2 Instructions to assemble a barbecue, from the box.



Figure 2.3 Participant assembling a TV cart using the picture on the box. For color version of this figure, the reader is referred to the online version of this book.

mental rotation task (Heiser, Phan, Agrawala, Tversky, & Hanrahan, 2004). The high spatial participants assembled the TV cart nearly twice as fast as the low spatial participants, with a fraction of the errors. The high spatial participants also produced far better visual instructions. Examples of each appear in Figure 2.4. Note that, despite clear differences, they have beginnings, segmented middles, and ends.

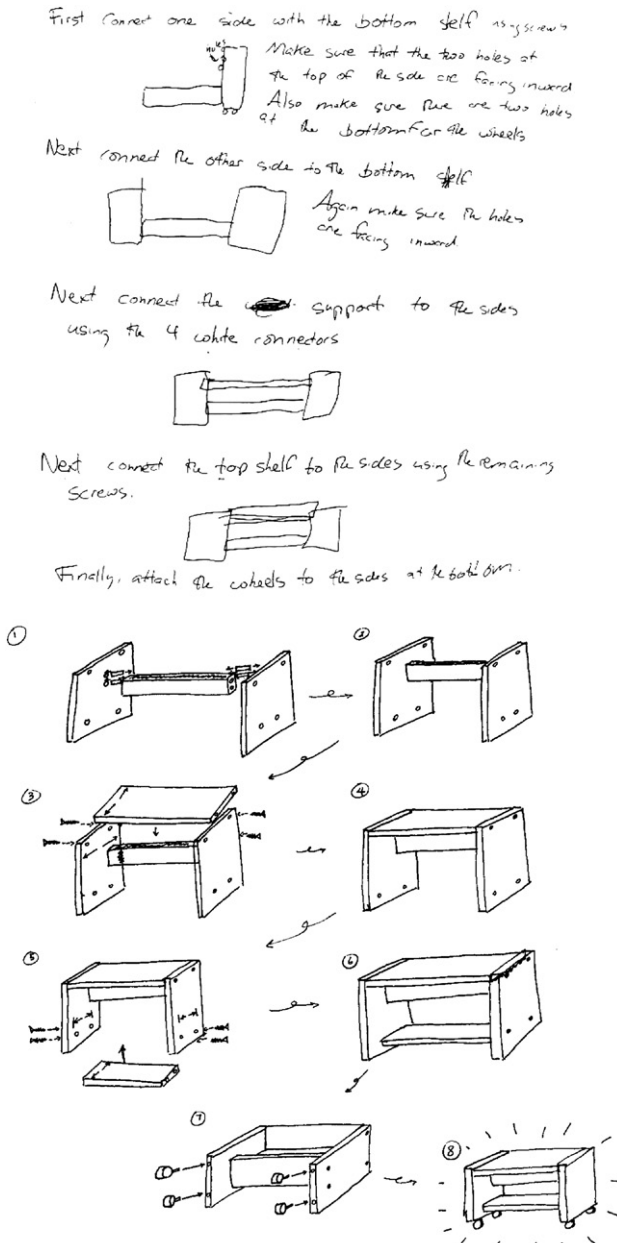


Figure 2.4 Participant-produced instructions to assemble the TV cart. The top drawing was produced by a low spatial participant and the bottom drawing was produced by a high spatial participant.

Importantly, diagrams of the high spatial participants showed the actions needed for assembly. They did this by selecting the viewpoint that showed the action, by using a 3-D perspective, and by adding non-depictive elements, guidelines, and arrows to show the actions needed to put the parts in the correct places. High spatial participants included three times as many action drawings as low spatial participants.

Yet another group assembled the TV cart and then rated the instructions produced by others. Participants, whether of high and low spatial ability, preferred the same diagrams: Those that were 3-D perspective drawings that showed the action step by step in the perspective of action. Reminiscent of the previous research on event segmentation, each new step introduced a new object part. Extracting from these studies yields three core cognitive design principles: show step by step, show perspective, show action.

The computer scientists down the street at Stanford then took over, developing an impressive algorithm that applied these principles (Figure 2.5). That algorithm began with a model of an object, then decomposed the object into integral parts, then decided on an assembly order that allowed effective assembly as well as effective diagrams. Finally, the algorithm produced step-by-step perspective action visual instructions. Lego instructions are regarded as the gold standard in the field; they can be used

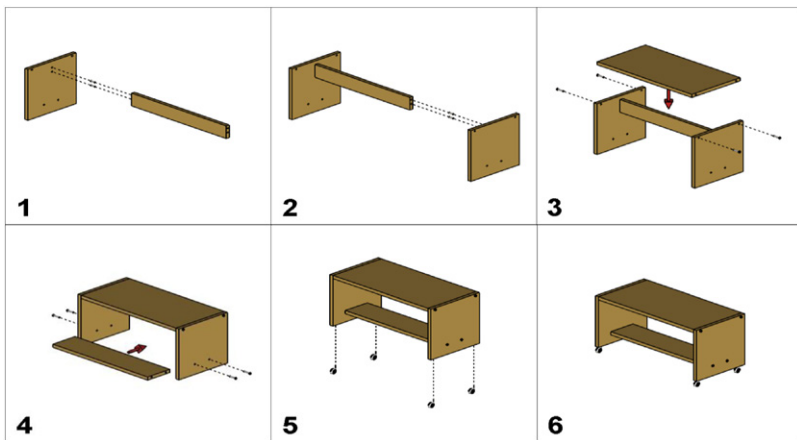


Figure 2.5 The instructions to assemble the TV cart produced by the computer algorithm based on the cognitive principles. (Adapted from Agrawala et al. (2003), p. 828 with permission). For color version of this figure, the reader is referred to the online version of this book.

by children all over the world. The algorithm recovered the standard Lego instructions from a completed Lego object. Here is what it produced for the TV cart:

Then came the judgment day. Would the beautiful instructions produced by the algorithm improve performance? Remember that participants could assemble the TV cart with no instructions at all. Remember that many studies comparing learning tools fail to show differences because people can learn even from mediocre explanations. So we ran a final (actually not, there were many other studies) study holding our breath. We recruited participants with low spatial ability, those perhaps in need of good instructions. Half were given the instructions from the algorithm. Half were given the instructions in the box, a part menu and exploded diagram, not as bad as the barbecue, but violating the cognitive design principles (see [Figure 2.6](#)).

We won! The participants were significantly faster, nearly twice as fast, and made fewer errors when using the step-by-step, perspective, action instructions than when using the instructions that came in the box. Whew!

Now some caveats. The cognitive design principles were not sufficient to produce the algorithm. A major discussion arose around adding the wheels, Step 5 above. The physics of the world advise turning the TV cart upside down to add the wheels. Should the diagram rotate the TV cart to show that? Showing a rotated TV cart might make it harder to recognize and confuse some users, especially those of low spatial ability. On the other hand, users might easily understand that inserting the wheels entails turning the cart. We could have done an experiment to decide, but instead, we went with our intuitions. In more complex situations, the relative costs of comprehending diagrammatic perspective switches or performing physical ones might differ, so changing the perspective of the drawings would probably be necessary, such as delicate attachments that need to be done from other perspectives. Similarly, scale changes may be needed, raising decisions about how to do that, globally, or with insets. Yet another key issue is amplifying visual instructions with language. Lego, our TV cart, and many other cases do not seem to need augmentation with words and symbols, but other cases are likely to benefit, and raise other design decisions: if words and symbols, which ones? For all these design decisions, whether to go with intuitions or to run experiments would depend on the relative costs and benefits of each.

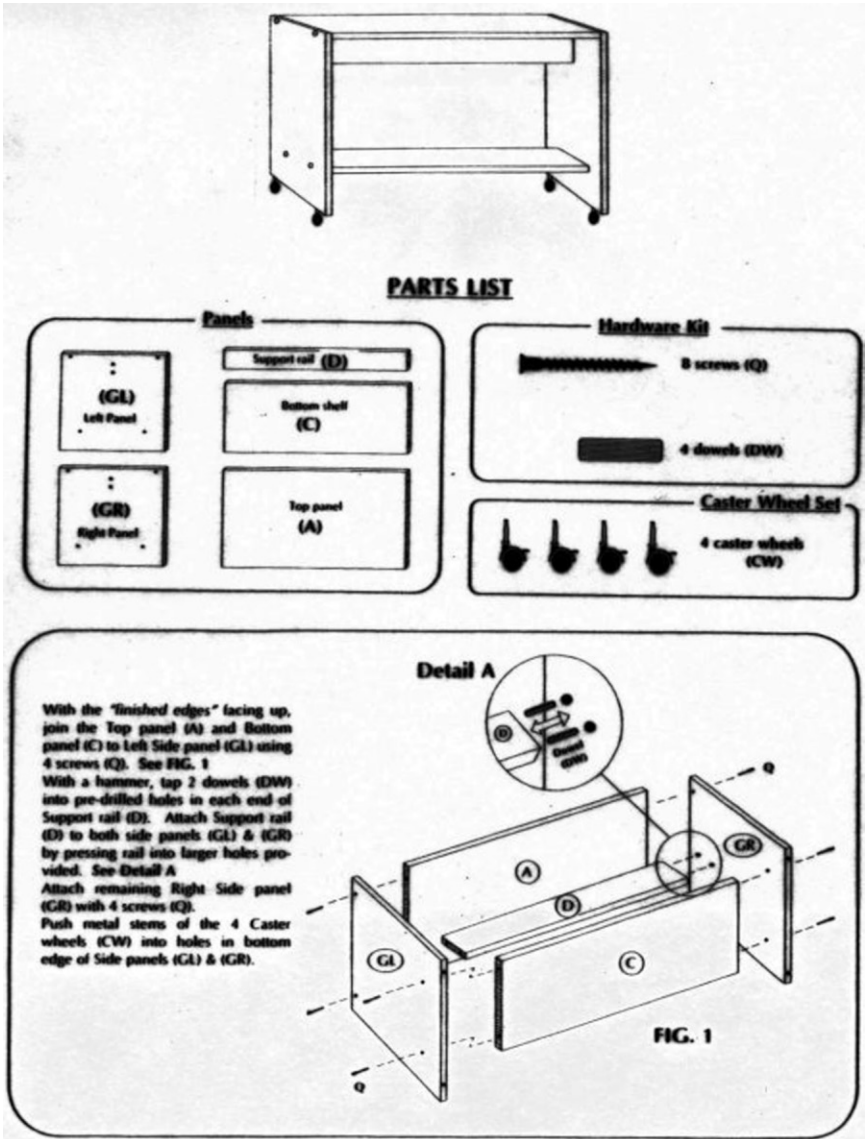


Figure 2.6 Instructions to assemble the TV cart, from the box.

More generally, design should be guided by empirically determined principles, but realistically, there is no time or money to try out every possibility. For these reasons, generally applicable design principles should be sought. We believe that the three principles uncovered

here—step by step, perspective, action—are such principles, and can be applied not only to the design of visual instructions but also to the design of visual explanations, how things work, and how to make things work.

This pair of successes, designing diagrams for routes and for assembly, led us to develop a general program for discovering cognitive design principles, based on three P's: production, preference, and performance. One group of participants with expertise produces diagrams for the desired end; another group expresses their preferences among the productions; a third group uses the diagrams to perform the target task. In the best of circumstances, the three P's converge to reveal a set of cognitive design principles that underlie the winning diagram (e.g. Kessell & Tversky, 2011; Tversky, Agrawala, et al., 2007; Tversky, Heiser, et al., 2007).

3.3. Visual Story Semantics: Depiction, Description, and Gesture

So far, the emphasis has been on visual stories. For the route maps, there was a small vocabulary, a toolbox of visual elements that turned out to be sufficient for participants to create many different routes (Tversky & Lee, 1999). Interestingly, these semantic elements had parallels in verbal elements. Paths were either straight lines or curved ones, corresponding to “go down” or “follow around.” Turns were T's, +’s, or L's corresponding to “turn” or “take a.” Although route maps could be analog, they were not, and they discretized the routes in the same ways, making the same distinctions, as non-analog language. Explanations using only gestures on maps made similar distinctions, notably, they pointed to landmarks and traced paths (Tversky, Heiser, Lee, & Daniel, 2009). There were parallels in the semantic elements used in depictions, descriptions, and gestures for explanations of TV cart assembly as well (Daniel & Tversky, 2012). The parallels suggest that the same modality-free mental representations, with their own semantics and syntax, underlie each mode of explanation. In spite of the parallels, some modes are bound to be more effective for some tasks than for others. Which modality for what information is a long discussion, and there is space for only some of that here.

Visual explanations, including both diagrams and gestures, have an advantage over purely symbolic ones: They can use elements in space and spatial relations to convey elements and relations that are literally or

metaphorically spatial. Maps illustrate the former and graphs the latter. There is a long list of other advantages for graphic representations, including allowing people's well-practiced skills in spatial reasoning to be used for abstract reasoning (e.g. Goldin-Meadow, 2003; Kendon, 2004; Kessell and Tversky, 2006; Kirsh, 1995; McNeill, 1992, 2005; Norman, 1993; Scaife and Rogers, 1996; Tversky, 1995; 2000b; 2001; 2004; 2005; 2011a; 2011b; 2011c; Tversky et al., 2003).

3.4. Principles of Apprehension and Congruence: Animation?

It goes without saying that any communication should be designed to conform to the Apprehension Principle, clear enough to be accurately perceived and comprehended. Mapping ideas to visual space raises another basic and general design principle: mappings to the page should be *congruent* with the desired thought (Tversky, Morrison, & Betrancourt, 2002). All other things being equal, larger quantities should take more space, related things should be closer than unrelated things, increases should go upwards. Purely symbolic language lacks spatial or visual congruence, but has one congruence fundamental to stories: sequence. Stories and explanations take place in time.

The Congruity Principle suggests that using time to convey events in time should have benefits. Novels, verbal or graphic, newspapers, plays, and films take advantage of that principle. But novels, newspapers, comics, plays, and films are highly crafted; they are not unedited presentations of life in real time. Notably, they segment action, often jumping from time to time and place to place, they change pace, they omit information, they interpret, they choose and change perspective, they often violate temporal order. Much as the mind does to actual experience. All too many animations meant to teach do not yet do the same.

Because animations meant to teach appear congruent and because technology has made creating them easier and easier, they have been adopted with enthusiasm. Surprisingly, a review of several dozen experiments failed to show benefits of animated over comparable static graphics (Tversky et al., 2002). The research included animations for children and for adults, for concrete physical concepts and for abstract algorithms, for explanations of how something works and explanations of how to do something.

In considering the many failures, it became clear that many did not conform either to the Principle of Apprehension or the Principle of Congruence. Too much happened too quickly to apprehend what, when, and

how, much less why. Most of the animations were continuous in time and merely showed a procedure or a process, they did not segment and they did not explain. However, as we have seen, people understand and explain the events of life as connected sequences of discrete actions, each ending in the accomplishment of a goal or subgoal. Routes are understood and explained as a sequence of turns at landmarks, assembly as a sequence of actions on objects. All segments are punctuated by completed actions, not by elapsed time. If the take-away message is a sequence of actions, then that is what should be presented, not an unbroken enactment in real time.

In fact, Julie Morrison and I (Morrison, 2000) found no advantage for animated graphics over static graphics when the task of participants was to learn a set of simple rules of motion for geometric figures. The movements were slow and simple, a ball rolling along a path, to avoid apprehension difficulties. Students learned the rules from a list or from static graphics, or from animated graphics. Below are the static presentations of two of the rules. In the animated version, a black ball moved along the indicated path. The gray areas represent barriers (Figure 2.7).

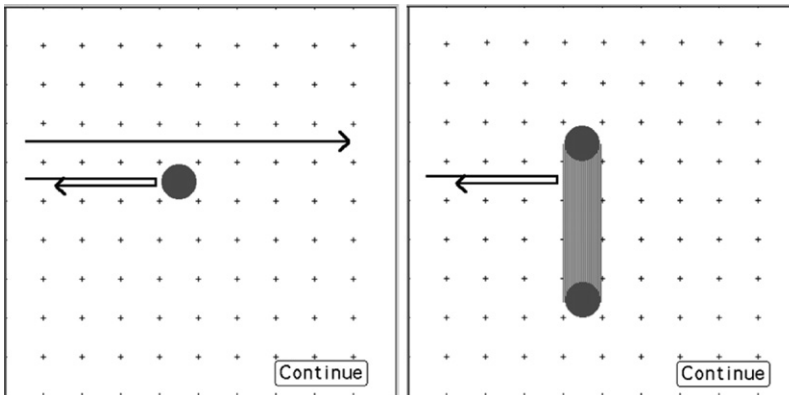


Figure 2.7 Static screen shot of some of the rules of movement.

In accordance with the Congruence Principle, both visual explanations were better than text alone, but there was no advantage to the animated visual explanation over the static one (Morrison, 2000). That is, viewing the permissible movements did not improve learning them.

Actual movement is fleeting. Static graphics remain in view where they can be inspected and reinspected, supporting apprehension. This suggests that there should be cases where animation is detrimental to learning, and it is. In experiments teaching students how to assemble complex toys,

animated graphics showing how to perform the specific attachments actually led to worse performance than static diagrams (Zacks & Tversky, 2003).

Some have proposed that interactivity can help overcome some of the apprehension difficulties. However, interactivity has not always yielded benefits. A set of experiments yoking participants to good or poor users of interactive animated graphics revealed that any benefits of the animations were due to intelligent selection of moments to view, not interactivity per se (Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008).

Clearly, there are effective animations. Films are a case in point—even without laboratory validation. Other cases are situations where continuous transitions help viewers keep oriented as things move or change in real time. For example, moving data points from one kind of representation to another, say scatter plots to bars, is helpful for keeping track of the changes (Heer & Robertson, 2007). Fill bars that keep us informed on the progress of some invisible process seem helpful. When the task is to recognize patterns of movement, such as the flocking of birds or the flow of liquids, there is no substitute for animation.

Designing animations that are effective for telling stories of processes or procedures that take place in time remains a challenge. Successful design requires a delicate balance of congruence—how the information is mapped—and apprehension—how the information is processed. As we said in a more extended analysis (Tversky, Heiser, et al., 2007): Seeing is not perceiving. Perceiving is not understanding. Showing is not explaining.

3.5. Static Tales: Word and Picture

Stories and explanations typically have structure, the social or physical relations among the characters or system parts, and action, the events that are driven at least in part by the social or physical structure. Structure, whether social or physical, is relatively easy to visualize, but visualizing action, events over time, is more difficult, and animations are not always effective. One way it's done is as a series of successive stills. In language, phrases, sentences, paragraphs, chapters, in the best cases, carefully crafted to clarify and motivate changes and transitions. In films, scenes. In graphics, successive frames, brought to an art form in comics (e.g. Eisner, 1985; McCloud, 1994). Successive frames, similar to successive phrases, sentences, paragraphs, require interpolation, bridging from one to another across myriad changes. For many graphic explanations and instructions, a single diagram is enriched with non-depictive devices that are meant to communicate change, notably arrows (e.g. Heiser & Tversky, 2006; van der Waarde &

Westendorp, 2000; Westendorp & van der Waarde, 2000, 2001). In diagrams of mechanical systems, arrows are reliably produced and interpreted as changes in time, action, and cause (Heiser & Tversky, 2006). One problem with arrows is exactly that: They have many possible meanings, change in time, change in action, motion, causality, invisible forces, and more, meanings that are all too often not disambiguated (e.g. Tversky, Agrawala, et al., 2007; Tversky, Heiser, et al., 2007).

3.6. Making Inferences

3.6.1. Designers Make Discoveries in Their Own Sketches

Essential as it is, making inferences from visual displays is often challenging. It is enhanced by expertise and ability (e.g. Anzai, 1991; Chase & Simon, 1973; Gobert, 1999; Koedinger & Anderson, 1990; Stylianou, 2002; Suwa & Tversky, 2001, 2003). Expert architects, for example, become adept at making unintended discoveries in their own sketches (e.g. Goldschmidt, 1994; Schon, 1983; Tversky & Suwa, 2009). Architects sketch their incomplete ideas for projects, and often see new unintended relations in their own sketches. Early sketches are typically ambiguous, and ambiguity encourages multiple interpretations. Expert architects are more likely to make functional inferences, such as traffic flow and diurnal and seasonal changes in lighting, than novices, who make primarily perceptual inferences, like finding patterns in configurations (Suwa & Tversky, 2001; Figure 2.8). Architects and designers report that reconfiguring or regrouping the elements of ambiguous sketches is a particularly helpful strategy for making discoveries and finding new interpretations. Suwa and I brought this task into the laboratory by asking designers and ordinary people to provide as many interpretations as they could for each of these four ambiguous sketches. For example, the second sketch from the left was often seen as a diagram or reconfigured as a robot; the third sketch was sometimes seen as an angel and sometimes reconfigured as a formation of rocks.

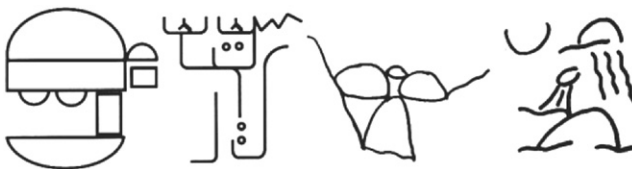


Figure 2.8 The four sketches used in the experiments of Suwa and Tversky (2001) and subsequent experiments. Used with permission.

Expert designers or architects were more adept at finding new interpretations than novices. Additionally, a perceptual ability, discerning simple figures in an intricate collection of complex ones (embedded figures), and a conceptual ability, making remote associations, independently promoted reinterpretations of sketches (Suwa & Tversky, 2003; Tversky & Suwa, 2009). Together, these abilities and associated processes underlie what we termed Perceptual Construction, the deliberate reconstruction of perception in the service of exploration and discovery of new ideas. It is a kind of design thinking that would seem to have generality to other domains of creative thought.

3.6.2. Making Inferences from Diagrams and Descriptions

Returning to explanations, Julie Heiser and I investigated the interplay of inferences from diagrams and expertise/ability. Structure, the spatial relations among the parts of the system, can be seen directly in diagrams, but function, the movements or changes in the parts as the system operates, must be inferred from diagrams. Arrows in diagrams encourage functional inferences (Heiser & Tversky, 2006), but they do not show the actual actions and changes. Thus, functional information must be inferred from diagrams, perhaps by mentally animating them (e.g. Hegarty & Just, 1993) but functional changes can be described directly in text. In our study, a total of 147 undergraduates studied one of two mechanical systems, a bicycle pump or a car brake. They were divided into high and low mechanical abilities on the basis of self-reports of general mechanical ability and knowledge of the system. The students learned by studying either a diagram that depicted the system or text that described the system. There were two kinds of diagrams and two kinds of text, structural and functional. The diagrams either depicted only the structure of the parts of the system, or they included arrows that indicated the actions of the system. Similarly, the text either described the structure of the system or described the actions of the system. Structural descriptions rely on *is-a* or *has-a* intransitive verbs, whereas functional descriptions rely on verbs of motion, typically transitive ones. This was the structural description for the car brake: The brake or brake drum is a circular structure. Directly inside the sides of the brake drum are two thick semicircular structures called the brake shoes. The brake fluid reservoir is located above and to the side of the brake drum. From the brake fluid reservoir, a tube runs down sideways and then down to the middle of the brake drum. Extending from both sides of the tube in the middle of the brake drum are wheel cylinders surrounding small pistons. Brake fluid can move from the reservoir through the tube to the pistons. The small pistons can

move outward toward the brake shoes. The brake shoes can move outward toward the brake drum. This was the functional description for the bicycle pump: When the handle is pulled up, it pulls the piston up. The pressure of the upward movement of the piston causes the inlet valve next to the piston at the top of the chamber to open and the outlet valve at the bottom of the chamber of the pump to close. This allows air to enter the lower chamber. When the handle is pushed down, pressure is exerted in the chamber causing the outlet valve to open. The pressure in the chamber and the opening of the outlet valve causes air to exit through the hose.

That design yielded eight groups, four learning conditions for each system. Students were later presented with true/false statements about the structure and function of the systems they had studied. Here is one structural statement about the car brake: The brake fluid reservoir is inside the brake drum. Now, a functional statement: The brake fluid pushes the brake drum outward. Accuracy and response times to verify the statements were recorded. Study time was self-determined and did not vary with any of the measures. Similarly, there were no effects on response times.

The largest effect was that of ability/expertise on accuracy for both structural and functional questions; those with high (professed) mechanical ability performed far better on both kinds of questions. There were eight of each. For structural questions, high ability/expertise participants made fewer errors ($M=1.59$, $SD=1.14$) than low ability/expertise participants ($M=2.5$, $SD=1.51$, $F[1, 139]=15.7$, $p<0.01$). For questions about function, high mechanical ability/expertise participants made fewer errors ($M=1.44$, $SD=1.3$) than low ability/expertise participants ($M=2.75$, $SD=1.6$, $F[1, 145]=29.6$, $p<0.01$).

For the questions about the structure of the system, as evident from [Figure 2.9](#), there were no effects of medium of instruction, diagrams and text were equally effective in instilling structural information. There were no interactions of medium of instruction and ability. The graph shows the errors made for each learning condition by each group, the white bars for the low ability/expertise group and the filled bars for the high ability/expertise group.

Performance on the functional questions yielded quite a different pattern, showing an intriguing interaction of learning format and ability/expertise, as evident in [Figure 2-10](#). As previously, there were eight functional questions, and the filled bars represent the high ability/expertise students and the white bars those of low ability/expertise. Medium, text or diagram, made no difference in conveying information about structure, but

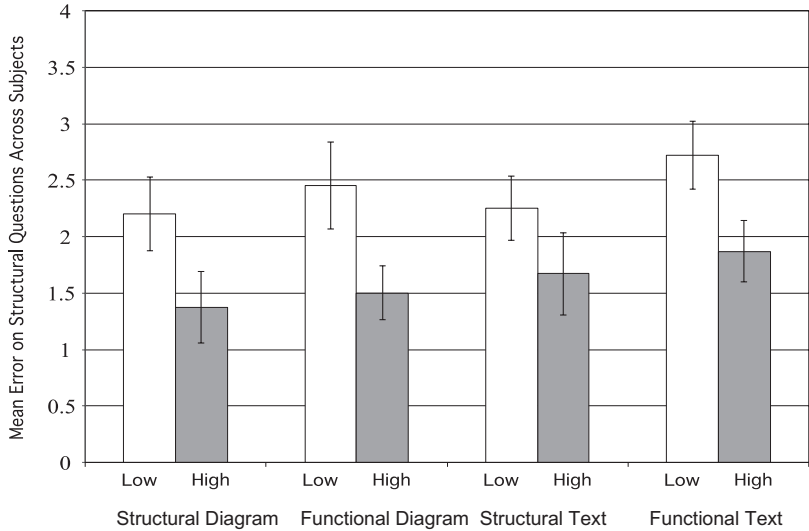


Figure 2.9 Errors on questions about structure by learning medium, structural or functional diagram or structural or functional text, and by mechanical ability, low or high.

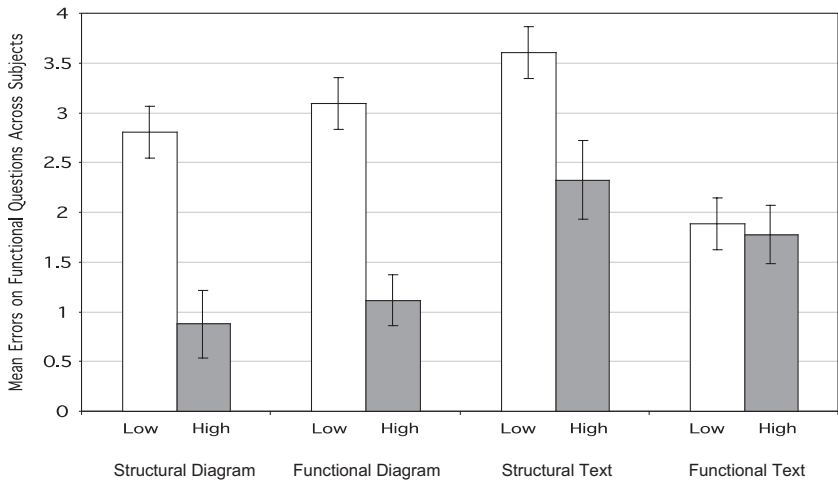


Figure 2.10 Errors on questions about function by learning medium, structural or functional diagram or structural or functional text, and by mechanical ability, low or high.

it had large effects in conveying information about function, effects that interacted with ability/expertise ($F[1, 139] = 8.02, p < 0.01$).

For high ability/expertise participants, those who studied diagrams made fewer errors on functional questions ($M = 1.1, SD = 1.1$) than those

who studied text, irrespective of the kind of diagram or text. By contrast, low ability/expertise participants who studied text made fewer errors on functional questions ($M=2.6$, $SD=1.6$) than those who studied diagrams ($M=3.0$, $SD=1.6$), equaling the performance of students of high ability/expertise. In addition, errors on functional questions were higher after studying a structural text ($M=3.0$, $SD=1.7$, [Figure 2.9](#)) than after studying a diagram without arrows ($M=1.87$, $SD=1.5$), functional text ($M=1.71$, $SD=1.27$, [Figure 2.10](#)) or diagram with arrows ($M=1.75$, $SD=1.68$, $F[1, 139]=17.48$, $p<0.01$).

The most striking finding was the effect of expertise/ability on acquiring functional information. Those high in expertise/ability extracted that information more easily from diagrams than from text. Those low in expertise/ability readily comprehended functional information from language but were less successful learning about function from diagrams. Put differently, when teaching is through diagrams, those of high ability/expertise have a distinct advantage over those of low ability/expertise in acquiring information about function; those low in ability/expertise make nearly three times as many errors. When teaching is from text that describes function, high and low ability/expertise participants perform equally well on questions about function. This finding jibes nicely with previous work showing that across many knowledge domains, people low in spatial or mechanical ability/expertise have difficulties making inferences from diagrams. Such results support an imperative for education urged by the [Committee on Support for Thinking Spatially](#), sponsored by the [National Academy of Sciences \(2006\)](#): It is essential to ensure that all students acquire expertise in comprehending and using diagrammatic information.

3.7. Gesture for Action

Just like life, stories and explanations in the wild are multimodal. They integrate words, prosody, gesture, actions, and props, including diagrams and sketches. Gestures serve many of the same roles, act in many of the same ways as diagrams. Both can use elements in space and spatial relations to convey spatial and abstract meanings. Both can use elements that carry resemblances to what they represent, that is, they can be iconic. Gestures cannot be as rich or refined as marks on paper, and for that reason, they are likely to be more abstract. They can show the proverbial big fish by a crude outline, but not by a detailed sketch. They can indicate an upward trend with a sweep of the hand, but not by an ordinate and abscissa marked with values. Gestures are necessarily more abstract, approximate, than depictions.

Depictions require tools, paper and pen. Gestures require only the tools we carry with us. Gestures, like words, disappear, but marks on paper stay there, to be contemplated on.

Gestures are actions, and, like animations, take place in time. As such, they are especially suited for conveying action. In a pair of studies, Kang, Tversky, and Black (2012; submitted) have found that explainers use gestures to convey action and that learners can acquire information about actions from gestures. In the study of explaining (Kang, et al, submitted), participants first learned complex systems, the circulatory system and the rock cycle, from diagrams. They then crafted explanations of the systems, enacting the role of a student by explaining the systems to imagined experts or enacting the role of a teacher by explaining the systems to imagined students. Participants typically used the given diagram in their explanations, but especially when explaining to novices, they also created larger diagrams in the air with gestures. Explainers gestured on both real and virtual diagrams. Many gestures, primarily deictic ones, conveyed the structure of the systems. However, the vast majority of gestures carrying semantic content conveyed the actions of the system. Those gestures were iconic gestures enacting the functions of the system.

In the study of learning (Kang, et al, 2012), two groups of participants watched a video of someone explaining the workings of a four-stroke engine. A large transparent diagram of the system was superimposed in front of the explainer. The explainer delivered the exact same verbal script but with two different sequences of gestures. For the function gesture sequence, the speaker enacted the actions of each step of the process, iconic gestures, without conveying any information about structure. For the structure gesture sequence, the explainer pointed to each part on the diagram in succession, deictic gestures, so no action information was given. A post-test that could be answered solely on the basis of the verbal script showed that all participants learned the basics of the structure and action of the systems. There were two transfer tasks: creating a visual explanation of the system and making a video to explain the system to a fellow student. Those who had seen gestures conveying action depicted more action in their visual explanations and used more gestures conveying action in their video explanations. Remember that it is action or function that must be inferred from diagrams and that is difficult for some to infer. Gestures in combination with diagrams appear to be an effective way to overcome those difficulties and to convey action effectively. Many gestures are actions, constituting a congruent way to convey action.

3.8. 'Real' Stories

Some of you are undoubtedly disappointed. How to get from here to there, how to put something together, how something works are not very interesting stories. Sure, it is important to learn those things, but these topics are hardly gripping, not the high drama of Shakespeare or the low (but sometimes more engaging) drama of gossip. In order not to disappoint, I turn now to work on that with Elizabeth Marsh and Nicole Dudokovic. At the beginning, I noted that life happens all at once from all directions, a multimodal light-and-sound show in space and time. Simply recounting what happened does not make a good story. There has to be a beginning, a middle, and an end; there has to be a point. The point directs the story, determining what is included, and what not, and how what is included is interlinked. Can the point then alter the memory of what happened? It seems that it can. In one of the experiments of [Tversky and Marsh \(2000\)](#), participants read dry straightforward descriptions of the activities of two new roommates throughout a college week. Each did prosocial things, like being the life of a party, and annoying things, like losing your new jacket, in equal numbers. After a break, the participants were asked to write a letter about one of the roommates either to recommend them to a fraternity or sorority or to request a change of roommate. In contrast to the original stories, the letters were lively and embellished with interpretations, selecting the prosocial activities of that roommate for the recommendation and the annoying activities for the request. A control group did not write a letter. Still later, all the participants were asked to recall the original descriptions they had read at the beginning of the session. Those descriptions were as dry and straightforward as the ones they had read. However, the descriptions of the letter writers were biased in the direction of the perspective, the point, of the letters they had written. Letter writers recalled more of the activities that were relevant to the letter's perspective. What's more, they contained intrusions relevant to the perspective. For example, if they had written to recommend David to a fraternity, but it was Michael who was the great dancer, participants often attributed that activity to David. Apparently, the perspective serves not only to organize and retrieve memory but also to reconstruct memory. Presumably, the weaker the memory for the actual activities, the stronger the effect of the organizing schema. Not just what happened, but what must have happened.

Different audiences and different motives elicit different spins, with consequent effects on memory. In another study ([Dudukovic, Marsh & Tversky, 2004](#)), participants read a humorous story about a novice

bartender's evening that contained both dramatic and boring events. Two days later, most of the participants were asked to retell the story to a video camera. Half of those were asked to retell the story as accurately as possible, essentially simulating eyewitnesses, and half were asked to retell the story in an entertaining way. The retellings were quite different, as expected. The entertaining retellings were longer and more fluent; they contained more affect, less perceptual information, and more frequently used the present tense. The retellings meant to be entertaining were rated as more entertaining but less accurate than the accurate retellings. Four days later, all the participants returned for recall and recognition tests of the original story. The recall by those who had given accounts meant to be accurate was higher than the recall by those who had given accounts meant to be entertaining. However, the recognition memory scores of the two groups did not differ, strengthening the claim that the spin participants use in retellings serves as a way of imposing coherence on a story and later acts as a scaffold for recall.

What happens in the wild? To address that, we asked students to keep track of the stories they told their friends and family each day (Marsh & Tversky, 2004). They recorded what, when, how, and to whom they related the events of their lives. They told stories differently depending on the audience. They exaggerated more when intending to entertain and simplified more when intending to inform. Although they reported distorting the content 61% of the time by exaggerating, omitting, minimizing, or embellishing their stories, they regarded only 42% of their stories as inaccurate. In other words, a certain degree of stretching the truth is seen as acceptable.

3.9. Explanations and Narratives

Both visual explanations and visual narratives—and for that matter, verbal ones as well—select segments of space and time and link them into coherent discourse. Participants break segments using top-down considerations, completion of goals, as well as bottom-up considerations, peaks of change in motion. Linking segments is accomplished by various forms of anaphora, including visual anaphora such as preserving an important visual element from one segment to the next, when continuity is desirable. When continuity is not desired, visual elements may be contrasted rather than preserved. Explanations of sequences of events enacted by people, like making a bed, are linked by accomplishments of goals and subgoals. Explanations of natural systems, like the way an engine works, are linked by actions and

outcomes. Both kinds of discourse link cause and effect, as do narratives, 'real stories.' What makes narrative distinct from explanations is that in narratives, there is a human voice, a point of view (e.g. Bruner, 1986). Typically, there is drama as well, a problem, then suspense, and finally, a resolution. The epitome of (static) visual narratives is comics, which use an impressive range of visuospatial poetic devices for conveying meaning, literally and metaphorically, as well as linking and breaking segments (see Eisner (1985) and McCloud (1994) as well as ongoing research with Jon Bresman). The uncertainty that drives that suspense is undoubtedly one of the reasons why 'real stories,' narratives are so compelling.



4. SPRACTION: SPACE, ACTION, ABSTRACTION

Now back to diagrams, gesture, and thought for the meta-story, the story we promised that will unite space, action, and abstraction.

People, all of us, design the world around us. We place books in rows on shelves, ordered by date or size or topic. We put dishes on shelves in kitchens, large plates in one pile, smaller ones in another, glasses and cups on other shelves, silverware sorted by types in drawers. We arrange one of each on tables for diners. Facades of buildings repeat windows and balconies for each room. Buildings are lined up along streets, interrupted by cross-streets, sidewalks, and street lamps. These actions, putting, placing, piling, lining, and distributing objects, create patterns that are good gestalts and easily recognized as the products of sentient minds. The patterns that are created reflect abstractions. Things are placed, piled, ordered, distributed for reasons, intriguing at least some observers to discern those reasons. The piles and rows and patterns carry abstract information. Archeologists, for example, use those regular arrangements as clues to the societies that created them. These patterns are present in time as well as space. The consequent arrangements in space and in time express abstractions: groupings, categories, hierarchies, orderings, dimensions, repetitions, symmetries, embeddings, and one-to-one or one-to-many correspondences. Those actions that create abstractions are the ways we organize and interrelate knowledge in our minds as well as objects and events in the world. The actions that create them are incorporated into gestures that express those abstractions in communications to ourselves and to others, gestures that both express and promote thought. These patterns that people create by their actions in space, rows and piles and lines and one-to-one correspondences and more, diagram the world. The same patterns are deliberately incorporated into diagrams

and graphs that express, indeed, communicate, the same abstractions: Space, Action, Abstraction, and Spraction.

It is difficult to talk about space and time because so many wise people have already said so much. It is difficult not to talk about space and time because space and time are everywhere always, an inextricable part of every event, the basis for thought, literal and figurative. Space and time pervade thought, and interconnecting space and time in patterns and explanations and stories is fundamental to being human.

ACKNOWLEDGMENTS

The author is grateful for many hours of great talk with colleagues and collaborators some of whose joint work is recounted here, including Jeff Zacks, Bridgette Martin Hard, Azadeh Jamalian, Seokmin Kang, Angela Kessell, Paul Lee, Marie-Paule Daniel, Masaki Suwa, Pat Hanrahan, Maneesh Agrawala, Julie Heiser, Julie Morrison, Mireille Betrancourt, Jane Nisselson, Andrea Kantrowitz, Elizabeth Marsh, and Nicole Dudokovic. Special appreciation goes to Julie Heiser and to Julie Morrison for granting inclusion of joint work in this article. The author is also indebted to the following grants for facilitating the research and/or preparing the manuscript: National Science Foundation HHC 0905417, IIS-0725223, IIS-0855995, and REC 0440103; the Stanford Regional Visualization and Analysis Center; and Office of Naval Research N00014-PP-1-O649, N000140110717; and N000140210534.

REFERENCES

- Agrawala, M., Phan, D., Heiser, J., Haymaker, J., Klingner, J., Hanrahan, P., et al. (2003). Designing effective step-by-step assembly instructions. *Proceedings of SIGGRAPH 2003* (pp. 828–837). ACM Press.
- Agrawala, M., & Stolte, C. (2001). Rendering effective route maps: Improving usability through generalization. *Proceedings of SIGGRAPH 2001* (pp. 241–250). ACM Press.
- Anzai, Y. (1991). Learning and use of representations for physics expertise. In K. A. Ericsson, & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 64–92). New York: Cambridge University Press.
- Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81.
- Committee on Support for Thinking Spatially. (2006). *Learning to think spatially*. Washington, DC: The National Academies Press.
- Daniel, M. -P., & Tversky, B. (2012). How to put things together. *Cognitive Processing*, 10.1007/s10339-012-0521-5.
- Denis, M. (1997). The description of routes: a cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, 16, 409–458.
- Denis, M., Pazzaglia, F., Cornoldi, C., & Bertolo, L. (1999). Spatial discourse and navigation: an analysis of route directions in the city of Venice. *Applied Cognitive Psychology*, 13, 145–174.
- Dudokovic, N., Marsh, E., & Tversky, B. (2004). Telling a story or telling it straight: the effects of entertaining versus accurate retellings on memory. *Applied Cognitive Psychology*, 18, 125–143.

- Eisner, W. (1985). *Comics and sequential art*. Tamarac, FL: Poorhouse Press.
- Gobert, J. D. (1999). Expertise in the comprehension of architectural plans. In J. Gero, & B. Tversky (Eds.), *Visual and spatial reasoning in design* (pp. 185–205). Sydney, Australia: Key Centre of Design Computing and Cognition.
- Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
- Goldschmidt, G. (1994). On visual design thinking: the vis kids of architecture. *Design Studies*, 15(2), 158–174.
- Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General*, <http://dx.doi.org/10.1037/a0024310>.
- Hard, B. M., Tversky, B., & Lang, D. (2006). Making sense of abstract events: building event schemas. *Memory and Cognition*, 34, 1221–1235.
- Heer, J., & Robertson, G. (2007). Animated transitions in statistical data graphics. *IEEE Transactions in Visualization and Computer Graphics. Proceedings of InfoVis*, 13, 1240–1247.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32, 717–742.
- Heiser, J., Phan, D., Agrawala, M., Tversky, B., & Hanrahan, P. (2004). Identification and validation of cognitive design principles for automated generation of assembly instructions. *Proceedings of advanced visual interfaces '04* (pp.311–319). ACM.
- Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, 30, 581–592.
- Kang, S., Tversky, B., & Black, J. B. (2012). From hands to minds: gestures promote action understanding. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 551–557). Austin, TX: Cognitive Science Society.
- Kang, S., Tversky, B., and Black, J. B. Gesture and speech in explanations to experts and novices, submitted for publication.
- Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: what matters is what you see, not whether you interact. *Cognitive Science*, 32, 1099–1132.
- Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.
- Kessell, A. M., & Tversky, B. (2006). Using gestures and diagrams to think and talk about insight problems. *Proceedings of the Meetings of the Cognitive Science Society*.
- Kessell, A. M., & Tversky, B. (2011). Visualizing space, time, and agents: production, performance, and preference. *Cognitive Processing*, 12, 43–52. <http://dx.doi.org/10.1007/s10339-010-0379-3>.
- Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, 73, 31–68.
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: elements of expertise in geometry. *Cognitive Science*, 14, 511–550.
- Marsh, E., & Tversky, B. (2004). Spinning the stories of our lives. *Applied Cognitive Psychology*, 18, 491–503.
- McCloud, S. (1994). *Understanding comics*. New York: Harper Collins.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental Brain Research*, 179, 427–442.
- Morrison, J. B. (2000). Does animation facilitate learning? An evaluation of the congruence and equivalence hypotheses. Unpublished PhD dissertation. Stanford University.
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28, 28–38. <http://dx.doi.org/10.1037/h0035584>.
- Newton, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12, 436–450. [http://dx.doi.org/10.1016/0022-1031\(76\)90076-7](http://dx.doi.org/10.1016/0022-1031(76)90076-7).

- Newton, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, 35, 847–862. <http://dx.doi.org/10.1037/0022-3514.35.12.847>.
- Norman, D. A. (1993). *Things that make us smart*. Reading, MA: Addison-Wesley.
- Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, 45, 185–213.
- Schon, D. A. (1983). *The reflective practitioner*. New York: Harper Collins.
- Stylianou, D. A. (2002). On the interaction of visualization and analysis: the negotiation of a visual representations in expert problem solving. *The Journal of Mathematical Behavior*, 21, 305–317.
- Suwa, M., & Tversky, B. (2001). Constructive perception in design. In J. S. Gero, & M. L. Maher (Eds.), *Computational and cognitive models of creative design V* (pp. 227–239). Sydney: University of Sydney.
- Suwa, M., & Tversky, B. (2003). Constructive perception: a skill for coordinating perception and conception. In *Proceedings of the cognitive science society meetings*.
- Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13, 407–433.
- Tversky, B. (1995). Cognitive origins of graphic conventions. In F. T. Marchese (Ed.), *Understanding images* (pp. 29–53). New York: Springer-Verlag.
- Tversky, B. (2000a). Remembering space. In E. Tulving, & F. I. M. Craik (Eds.), *Handbook of memory* (pp. 363–378). New York: Oxford University Press.
- Tversky, B. (2000b). Some ways that maps and graphs communicate. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Spatial cognition II: Integrating abstract theories, empirical studies, formal methods, and practical applications* (pp. 72–79). New York: Springer.
- Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 79–111). Cambridge: MIT Press.
- Tversky, B. (2004). Semantics, syntax, and pragmatics of graphics. In K. Holmqvist, & Y. Ericsson (Eds.), *Language and visualisation* (pp. 141–158). Lund: Lund University Press.
- Tversky, B. (2005). Functional significance of visuospatial representations. In P. Shah, & A. Miyake (Eds.), *Handbook of higher-level visuospatial thinking* (pp. 1–34). Cambridge: Cambridge University Press.
- Tversky, B. (2011a). Spatial thought, social thought. In T. Schubert, & A. Maass (Eds.), *Spatial schemas in social thought*. Berlin: Mouton de Gruyter.
- Tversky, B. (2011b). Tools for thought. In B. Benedetti, & V. Cook (Eds.), *Language and bilingual cognition* (pp. 131–139). New York: Psychology Press.
- Tversky, B. (2011c). Visualizations of thought. *Topics in Cognitive Science*, 3, 499–535.
- Tversky, B., Agrawala, M., Heiser, J., Lee, P. U., Hanrahan, P., Phan, D., et al. (2007). Cognitive design principles for generating visualizations. In G. Allen (Ed.), *Applied spatial cognition: From research to cognitive technology* (pp. 53–73). Mahwah, NJ: Erlbaum.
- Tversky, B., Heiser, J., Lee, P., & Daniel, M.-P. (2009). Explanations in gesture, diagram, and word. In K. R. Coventry, T. Tenbrink, & J. A. Bateman (Eds.), *Spatial language and dialogue* (pp. 119–131). Oxford: Oxford University Press.
- Tversky, B., Heiser, J., Lozano, S., MacKenzie, R., & Morrison, J. (2007). Enriching animations. In R. Lowe, & W. Schnotz (Eds.), *Learning with animation*. Cambridge: Cambridge University Press.
- Tversky, B., & Lee, P. U. (1998). How space structures language. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition: an interdisciplinary approach to representation and processing of spatial knowledge* (pp. 157–175). Berlin: Springer-Verlag.
- Tversky, B., & Lee, P. U. (1999). Pictorial and verbal tools for conveying routes. In C. Freksa, & D. M. Mark (Eds.), *Spatial information theory: cognitive and computational foundations of geographic information science* (pp. 51–64). Berlin: Springer.
- Tversky, B., & Marsh, E. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology*, 40, 1–38.
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human Computer Studies*, 57, 247–262.

- Tversky, B., & Suwa, M. (2009). Thinking with sketches. In A. Markman (Ed.), *Tools for innovation*. Oxford: Oxford University Press.
- Tversky, B., Suwa, M., Agrawala, M., Heiser, J., Stolte, C., Hanrahan, P., et al. (2003). Sketches for design and design of sketches. In Ugo Lindemann (Ed.), *Human behavior in design: Individuals, teams, tools* (pp. 79–86). Berlin: Springer.
- Tversky, B. & Zacks, J. M. Event perception. In D. Riesberg (Ed.), *Oxford handbook of cognitive psychology*. Oxford: Oxford, in press.
- Tversky, B., Zacks, J. M., & Hard, B. M. (2008). The structure of experience. In T. Shipley, & J. M. Zacks (Eds.), *Understanding events* (pp. 436–464). Oxford: Oxford University.
- Tversky, B., Zacks, J. M., Morrison, J. B., & Hard, B. M. (2010). Talking about events. In E. Pederson, J. Bohnemeyer, & R. Tomlin (Eds.), *Event representation in language and cognition* (pp. 216–227). Cambridge: Cambridge University Press.
- van der Waarde, K., & Westendorp, P. (2000). The functions of arrows in user instructions. Unpublished manuscript.
- Westendorp, P., & van der Waarde, K. (2000/2001). Icons: support or substitute? *Information Design Journal*, 10, 91–94.
- Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127, 3–21.
- Zacks, J. M., & Tversky, B. (2003). Structuring information interfaces for procedural learning. *Journal of Experimental Psychology: Applied*, 9, 88–100.
- Zacks, J., Tversky, B., & Iyer, G. (2001). Perceiving, remembering and communicating structure in events. *Journal of Experimental Psychology: General*, 136, 29–58.



The Cognition of Spatial Cognition: Domain-General within Domain-specific

Holly A. Taylor*,¹, Tad T. Brunyé*,**

*Department of Psychology, Tufts University, Medford, MA, USA

**U.S. Army Natick Soldier Research, Development, and Engineering Center, Natick, MA, USA

¹Corresponding author: E-mail: holly.taylor@tufts.edu

Contents

1. Spatial Cognition: Exploring the Domain-general in the Domain-specific	78
2. Spatial Cognition and Attention	79
2.1. Binding What and Where: Implications for Automaticity	80
2.2. Goal Processing and Attention	84
2.3. Affect and Arousal: General Implications for Spatial Cognition	87
2.4. Technology for and During Navigation: Attention Distractors	89
2.5. Conclusions: Attention and Spatial Cognition	91
3. Memory and Spatial Cognition	91
3.1. Working Memory Involved in Spatial Cognition	91
3.2. Structuring Spatial Memory	93
3.3. Conclusions: Memory and Spatial Cognition	96
4. Representing Spatial Information: Embodiment	96
4.1. Embodiment: Spatial Reference Frames and Representational Perspective	101
4.1.1. <i>Embodiment and Use of Alternative Reference Frames</i>	102
4.1.2. <i>Embodiment of Route and Survey Perspectives</i>	104
4.2. Embodied Spatial Representations: Conclusions	107
5. The Cognition of Spatial Cognition	108

Abstract

Few would argue against the position that spatial cognition involves cognition. Much of spatial cognition research has focused on illuminating the domain-general processes (e.g. attention, memory, or representation) active in this domain-specific field. In this chapter, we suggest that researchers view this domain-general to domain-specific relationship in the opposite direction. In other words, we review spatial cognition research within the context of its utility for understanding domain-general processes. For a cognitive process to be domain-general, it should be evident across a wide variety of domain-specific tasks, including verbal and spatial ones. Yet, the majority of data supporting domain-general models comes from verbal tasks, such as list

learning. Thus, we suggest that considering spatial cognition tasks and findings along with those from other domains would enhance our understanding of truly domain-general processing.



1. SPATIAL COGNITION: EXPLORING THE DOMAIN-GENERAL IN THE DOMAIN-SPECIFIC

Spatial cognition as a subdiscipline of cognition falls under a category referred to as *domain-specific cognition*. This category has an isolating effect. It suggests work that is narrowly focused, thus carrying implications for how the larger field of Cognitive Science views spatial cognition research. The outcome of this narrow view is what happens in spatial cognition research stays within spatial cognition research. Spatial cognition citations appear in other spatial cognition studies, but rarely venture into either domain-general or other domain-specific work. This makes the field relatively insular.

In reality, all cognitive research, whether conceptualized as domain general or domain specific, has domain-specific elements. For researchers to study cognition, they must have participants engage in cognition. In research studies, participants take on a researcher-generated task, a task embedded in some domain. For example, to study memory, participants need to encode, store, and retrieve information from memory. In much of basic memory research, participants learn and then remember lists, sometimes words (e.g. the false memory literature has a huge base in the Deese-Roediger-McDermott (DRM) list-learning paradigm; [Roediger & McDermott, 1995](#)). Results of these studies are then discussed in a domain-general framework of “memory processes” and are used in developing and testing general memory models. Little, if anything, is mentioned about the nature of the task, the incorporated semantic meaning of words, and/or other domain-specific task features. Yet we know that task differences, instruction differences, and even stimuli differences can affect task performance. Thus, tasks introduce a domain. But, for a cognitive process to be domain-general, it should be evident across a wide variety of domain-specific tasks, including verbal and spatial ones. We suggest that spatial tasks, when considered in concert with other tasks varying in modality and content, can contribute to domain-general understanding of cognitive processes.

This chapter aims to review spatial cognition research, tying it to a domain-general framework. In particular, we discuss how spatial cognition findings explicate a range of domain-general processes. We focus on three cognitive processes in particular: attention, memory, and representation. We further argue

for the potential of bidirectional benefits between domain-general and domain-specific understanding. In other words, a careful examination of the spatial cognition literature may enhance domain-general knowledge just as thinking about domain-general processes contributes to understanding spatial cognition.



2. SPATIAL COGNITION AND ATTENTION

Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence.

William James (1890; pp. 403–404).

Attention underlies both basic (e.g. perception and memory) and higher-order cognition, including spatial cognition. Without attention, information receives minimal, if any, processing. Two everyday spatial cognition experiences illustrate this fact. Passengers remember less about an environment than drivers do (Walmsley & Jenkins, 1991), and people have poor memory of environments traversed using a navigational aid (e.g. Gardony, Brunyé, Mahoney, & Taylor, 2012a). Although these outcomes involve more than attention, they are most commonly explained by reduced attention to the environment (e.g. Burnett, 2000). Spatial attention deficits also underlie two spatial cognitive neurological disorders, visual neglect and simultanagnosia (e.g. Dalrymple, Birmingham, Bischof, Barton, & Kingstone, 2010; Halligan, Fink, Marshall, & Vallar, 2003).

Basic attention research uses carefully controlled experimental paradigms. From these studies have come important advances in understanding attention. Within these studies, stimulus types have varied, including letters (Shapiro, Raymond, & Arnell, 1997), words (masked and unmasked; Lachter, Forster, & Ruthruff, 2004), and spatial information such as arrows (e.g. flankers task; Eriksen, 1995), auditory information (e.g. Moray, 1959), and visual scenes (Rensink, O'Regan, & Clark, 1997), to name some. While researchers have discussed task and stimuli variation used for attention research, these differences are largely glossed over when theorizing about attention in general.

In studying attention in situ, such as with higher-order cognition, attention, perception, and memory play interdependent roles (Chun & Turk-Browne, 2007). This interdependence makes sense. Attention enables us to select behaviorally relevant information for processing while ignoring

less relevant information. In other words, higher-order processes build on multiple basic processes. Although the interdependence is recognized, for simplicity, researchers largely explore these three cognitive subdomains separately from one another (Awth & Jonides, 2001; Chun & Turk-Browne, 2007). Yet, understanding the role of attention in higher-order cognition seems important to understanding attention itself. Several lines of spatial cognition research provide suggestions for the role attention plays in a real-world context.

2.1. Binding What and Where: Implications for Automaticity

Although few studies have explored how attention, perception, and memory interact in real-world tasks, their interdependence seems intuitive. Lack of attention or attention focused elsewhere leads to a cascade of cognitive consequences. People fail to perceive even unusual and obvious events and objects when they focus attention elsewhere (Hyman, Boss, Wise, McKenzie, & Caggiano, 2010; Simons & Chabris, 1999; Strayer & Drews, 2007), a phenomena known as *inattention blindness*. In the now modern classic, participants miss seeing a gorilla dancing through a basketball drill when tasked with counting the number of passes between teammates (Simons & Chabris, 1999). If something is not perceived, it will not be encoded in memory. However, some processes require more resources than others do, as seen in differences between automatic and controlled processing (e.g. Spelke, Hirst, & Neisser, 1976).

Cognitive automaticity has been explored with spatial information. Knowing a location means knowing where a “specific something” is located. This means attending to an object’s identity and its location (Ungerleider & Haxby, 1994), and putting these together. Neurally, neuroanatomically distinct parallel functional pathways process identity (*what*) and spatial (*where*) information (Postma, Kessels, & Van Asselen, 2008; Smith & Jonides, 1997). Thus, processing spatial information involves binding identity and spatial location, which involves attention to both. Attention directed elsewhere would affect this binding and the resultant memory. While binding different features clearly requires attention (Hyun, Woodman, & Luck, 2009; Treisman, 1996; Treisman & Gelade, 1980), attentional demands may not be equivalent on the features individually. Notably, some researchers have argued that location information is processed and remembered automatically (e.g. Hasher & Zacks, 1978). Automatic processing requires fewer attentional resources than does controlled processing (e.g. Shiffrin & Schneider, 1977). Hasher and Zacks (1978) functionally defined automatic

processing as requiring minimal attentional demands, co-occurring (but without interfering) with other cognitive processes, and operating equivalently with intentional and incidental learning. Together, these requirements predict that instruction and practice should have little, if any, effect on automatic processes. Automatic processes should also develop earlier, be largely unaffected by aging, and exhibit little individual variability.

Numerous studies support location processing as automatic (e.g. Ellis, Katz, & Williams, 1987; Mandler, Seegmiller, & Day, 1977; Von Wright, Gebhard, & Karttunen 1975), discussing it as akin to preattentive processing (Treisman & Gelade, 1980). Exploring intentional and incidental processing, Mandler et al. (1977) found that children and young adults recalled object locations equivalently under both incidental (attend to object identity) and intentional (attend to location and identity) instructions. Von Wright et al. (1975) also found no differences between incidental and intentional learning in memory for picture locations. Further, specific instructions to attend to picture locations did not benefit recall. Meeting the attentional demands and co-occurrence definition, Postma and De Haan (1996) showed that location errors did not increase with display size (i.e. number of items) increases or when participants engaged in a secondary task. This finding is consistent with Treisman and Gelade's (1980) preattentive feature processing. Looking at the developmental components of automaticity, Ellis et al. (1987) found little age and intelligence variability in location memory. All participants (children, mentally challenged adults, younger adults, and older adults), except the youngest children (ages 3–4), recalled picture locations equivalently well.

Other research, while not meeting with Hasher and Zacks' (1978) automaticity definitions, finds superior location memory. Johnston and Pashler (1990) explored identity and location binding in feature perception. They found close binding of identity and location, but an asymmetry between feature and location identification. Specifically, they found no perception of identity without location, but a trend for accurate location perception without identity. Thomas, Bonura, Taylor, and Brunyé (2012) examined visuospatial working memory in younger and older adults. Participants viewed 5×5 grids showing two to five objects (e.g. Figure 3.1). Memory for the grid was tested using one of three "yes–no" recognition tests. *Identity* memory trials showed an object and asked participants whether the object had appeared in the previously studied grid. *Location* memory trials showed an empty grid with one cell highlighted and asked whether something had appeared in that location. *Location/identity binding* trials showed an object

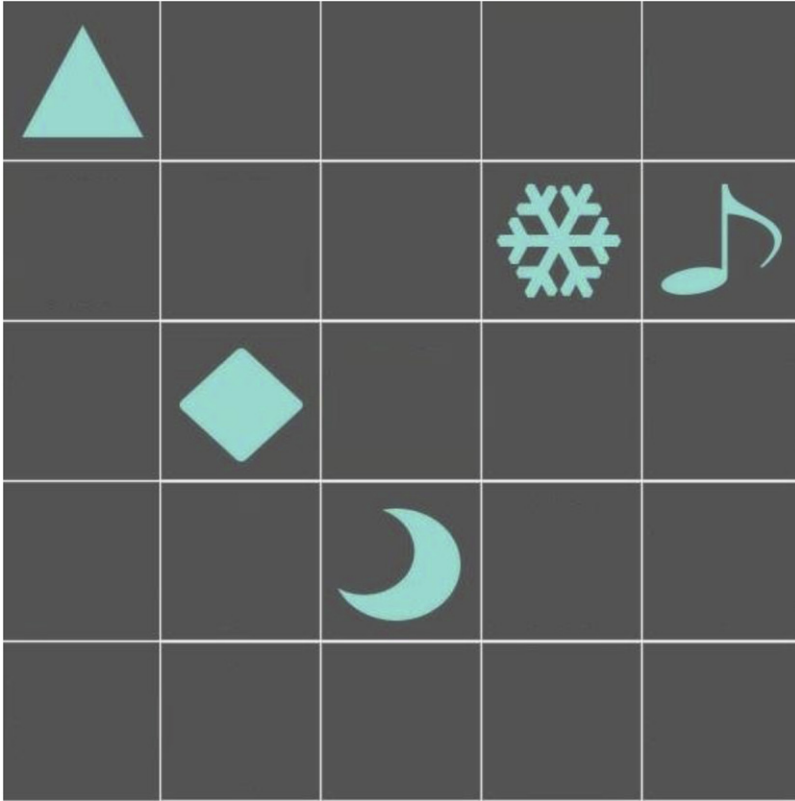


Figure 3.1 Sample stimuli from the study of [Thomas, Bonura, Taylor, et al. \(2012\)](#) examining binding of object identity and location. For color version of this figure, the reader is referred to the online version of this book.

in a grid cell and asked whether that particular object had appeared in that particular location. Experiment 1 blocked recognition trial types, allowing for intentional processing; experiment 2 randomized the trial types. Supporting less effortful location processing, location memory remained stable as the number of grid objects increased. This pattern was evident with both intentional (experiment 1) and incidental (experiment 2) processing. In contrast, identity and location/identity binding memory decreased with more objects. Inconsistent with automatic processing, older adults had worse location recognition than did younger adults. These age differences could be ameliorated when older adults had additional study time. Thus, location processing appears to require less attention, but does not meet the qualifications of an automatic process.

Different explanations come to mind for why spatial location might be processed with greater ease. One explanation lies in spatial control of attention. Observers can direct visual processing toward specific locations (Awh & Jonides, 2001). Information at these locations is then processed better and faster. Awh and Jonides (2001) argued that directing or biasing processing to specific locations in this way directly serves spatial working memory. They refer to this as *spatial rehearsal effects*. They base their argument on the fact that spatial working memory and spatial selective attention recruit largely overlapping, right-hemisphere neural networks centered in frontal and parietal areas. Behavioral evidence of spatial rehearsal effects can be seen in better memory for attended locations and reduced memory accuracy with attentional disruptions to specific locations. Important to this chapter is the fact that spatial direction of attention is not limited to spatial cognitive tasks; it plays a key role in other domains, such as reading (Rayner, 1979).

Another explanation lies in the role of location for guiding actions. This explanation arises from *just-in-time* models explaining perceptual representations (e.g. Ballard, Hayhoe, Pook, & Rao, 1997) and has implications for a wide variety of behaviors. *Just-in-time* models have been applied to *change blindness* (Simons, 2000). Change blindness describes the inability to see changes in objects or scenes across perceptual instances (e.g. blinks, saccades). Although paradigmatically different than studies examining binding of identity and location information, detecting changes across scenes requires identity/location binding. *Just-in-time* models accept the fact that successful action within an environment requires that a representation of some information be maintained. Given limitations on cognitive processing, maintaining location, or layout information may better serve future actions (Simons, 1996; Wang & Simons, 1998). Specifically, if one knows where something exists, but does not remember what it is, they know where to direct attention for further processing. Strong support for this contention about automaticity comes predominantly from tasks demanding egocentric (vs. allocentric) processing, which is generally more reflective of directly perceived viewpoints during guided action (cf. Pouliot & Gagnon, 2005).

The automaticity, or even superiority, of spatial processing, however, is not universally supported (e.g. Light & Zelinski, 1983; Park, Puglisi, & Lutz, 1982; Schulman, 1973). Contradictory evidence suggests that location information is better remembered when intentionally studied. Supporting evidence shows that participants locate map landmarks better after attending to both landmark identity and location than to just landmark identity alone (Light & Zelinski, 1983; Schulman, 1973). Research has also shown

notable age-related declines in location memory (e.g. Light & Zelinski, 1983). Chalfonte and Johnson (1996) found age-related declines in identity, location, and combined identity/location memory. Kessels, Hobbel, and Postma (2007) suggest that declines in location (context) and combined memory may underlie general age-related declines in episodic memory.

Thinking about location as “context” brings these results into a more general discussion of context on memory. Spatial location serves as one context, but semantic information can also provide context and one that has frequently been used in memory research. Reinstatement of context generally aids memory, as seen in *transfer appropriate processing* (Franks, Bilbrey, Lien, & McNamara, 2000). However, a debate continues on whether aging affects use of context (Spencer & Raz, 1995; Thomas & Bulevich, 2006). Spencer and Raz (1995), based on a meta-analysis, suggested a greater impact of age on context than on content memory. Specifically, the greatest age difference appeared when contextual features could be encoded independently from content. This has particular implications for spatial cognition. The separate *what* and *where* processing streams strongly suggest that spatial and identity information can be encoded separately. In a real-world context, this makes sense. If people perceive object locations as transient, for example, that one’s car keys can often be found in different locations or that the restaurant on the corner keeps changing ownership and name, they more likely encode identity (content) and location (context) separately. In contrast, Thomas and Bulevich (2006) suggest that older adults encode contextual cues but may have difficulty using them. Thomas, Bonura, and Taylor (2012) found similar results with older adults’ map learning.

The debate about whether spatial information is processed automatically, or more effortlessly, ties into several other cognitive research questions. It clearly has implications for general questions of automatic versus controlled processing. In addition, because spatial location provides context, this debate can inform research on context effects more generally. Finally, it has implications for how directing attention affects outcomes of a variety of tasks.

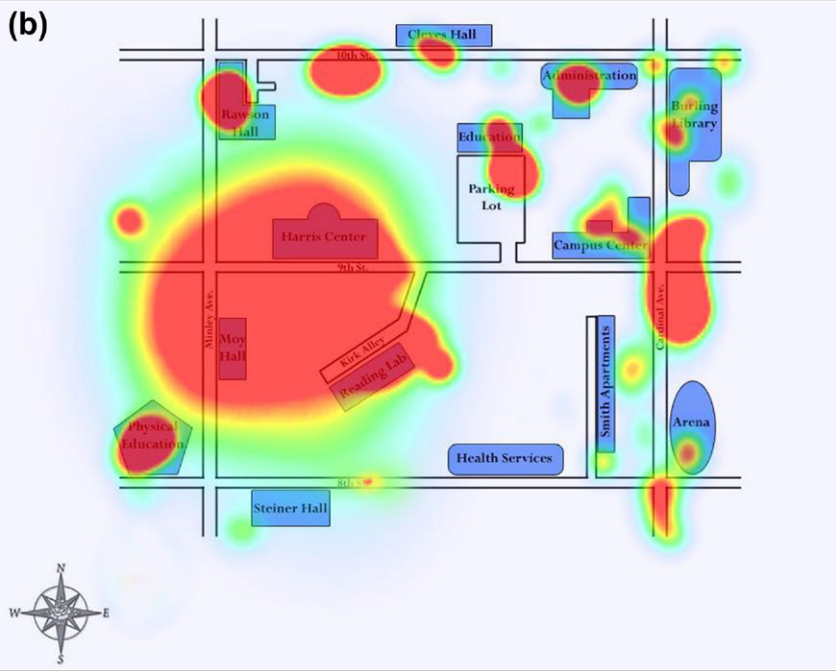
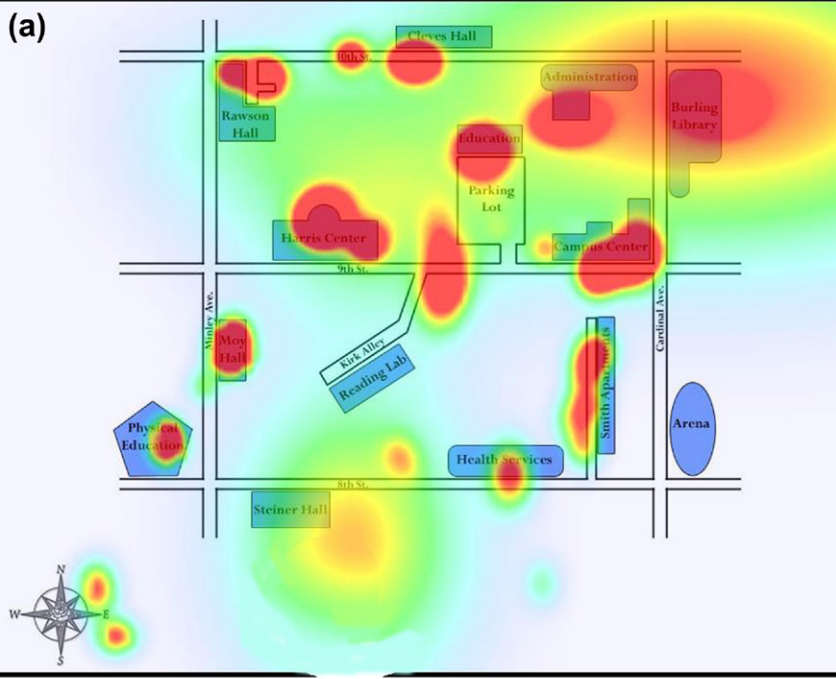
2.2. Goal Processing and Attention

We remember spatial locations for a purpose. That purpose may be as simple as finding our keys when we once again need to leave the house. For memory of larger-scale spaces, our purposes may be diverse. When learning a larger-scale environment, one generally intends to use that information for a specific purpose. In other words, they have a goal. If you are attending a weekend conference and need to know how to get from the airport

to your hotel, you may only gather route-relevant information from a map. Alternatively, if you have moved to a new city and want to get the “lay of the land,” you may take away information about salient landmarks and major thoroughfares. Goals alter where we direct our attention (Hopfinger, Buonocore, & Magnun, 2000; LaBerge, 1995; Maruff, Danckert, Camplin, & Currie, 1999), which in turn alters what is mentally represented and/or retrieved (Lichtenstein & Brewer, 1980). Showing evidence of this view, Pichert and Anderson (1977) had people read a story about a house either from the perspective of someone wishing to buy or rob the house. People remembered story details consistent with their goal perspective. When later asked to recall the story again, but taking the alternative perspective, people recalled some new details, consistent with their new perspective. In other words, behavioral goals can direct attention both during encoding and retrieval.

Behavioral goals change how people process information. Brunyé and Taylor (2009) examined eye movements during map study under three goal conditions. Participants studied the same campus maps, instructed either to learn the layout as it relates to canonical directions (survey goal), to learn the routes and landmarks along the routes (route goal), or to learn everything they could (unspecified goal). Results showed goal-consistent eye movement during initial map study. During the first few minutes of study, survey goal participants alternated focus between buildings and the compass rose. In contrast, route goal participants maintained focus on streets and street names (see Figure 3.2 for eye-movement heat maps). The fact that these goal-consistent eye movements only appeared early during study suggests that goal-related attentional influences help build the initial framework to which later study is then related. The memory results from this study support this contention. Participants were better able to verify spatial statements providing information consistent with their goal perspective, suggesting the use of a goal-consistent framework for retrieval. Gauvain and Rogoff (1986) also showed that goals guided study strategies. Children given the goal of learning the layout of a funhouse spent more time atop a slide that afforded an overview of the layout.

Goals, however, seem to be guides, but not blinders. In the study of Pichert and Anderson (1977), readers seemed to encode information unrelated to their study goal, although much of this information was not initially recalled. This is further evidenced in the fact that only initial eye movements in Brunyé and Taylor (2009) were more frequently focused on goal-consistent information. Additional behavioral evidence comes from Taylor,



Naylor, and Chechile (1999). In their work, participants either navigated a complex, unfamiliar building or studied a map. They learned through their assigned method with one of two goals, to learn the overall layout (survey goal) or the routes between rooms (route goal). Participants then completed several memory tasks, some consistent with a map/survey perspective and others consistent with navigation/route perspective. Results showed that learning method and learning goal interacted. Specifically, map study led to better survey perspective task performance, but a survey goal during navigation boosted survey performance for navigators. Likewise, navigators performed better on-route perspective tasks, but a route goal boosted route task performance for those who studied maps.

In summary, to understand what information people will represent and how they might use it, it is important to know how they intend to use the information. Goals guide information gathering strategies (Brunyé & Taylor, 2009; Gauvain & Rogoff, 1986), directing attention to goal-relevant information. Goal-relevant information processed early in learning further instantiates schemas that guide further study (Hopfinger et al., 2000; LaBerge, 1995; Maruff et al., 1999). But, these findings are not limited to spatial information. Van den Broek, Lorch, Linderholm, and Gustafson (2001) examined readers' goals (information or entertainment) on inference generation and memory for expository text. Goals affected reading strategies, and as a consequence, inferences. Information gathering goals yielded strategies to build coherence while entertainment goals yielded more associations. These reading strategies then carried over to memory. Thus, work exploring goal-directed spatial learning has implications for general goal-directed cognition.

2.3. Affect and Arousal: General Implications for Spatial Cognition

Emotions, including valence and arousal states, influence cognitive processing, including attention (Clore & Huntsinger, 2007; Gasper & Clore, 2002). The *levels-of-focus hypothesis* predicts that positive affective cues promote relational processing and negative cues promote item-specific processing (Clore et al., 2001). The levels-of-focus hypothesis has implications for the



Figure 3.2 Eye-movement heat maps showing differential eye-movement patterns during a study as a function of having a survey (a) or route (b) perspective goal. Heat maps based on data from Brunyé and Taylor (2009). For color version of this figure, the reader is referred to the online version of this book.

hierarchical structure of spatial information (e.g. Maddox, Rapp, Brion, & Taylor, 2008; McNamara, Hardy, & Hirtle, 1989; McNamara, Ratcliff, & McKoon, 1984; Stevens & Coupe, 1978). With this structure, one can focus attention on global or local levels (e.g. Navon, 1977). A complete understanding of spatial cognition then requires understanding factors implicated in shifting attention between local and global processing, including emotion. Comprehensive knowledge of both close (local) and far (global) landmark locations critically underlies successful navigation (Foo, Warren, Duchon, & Tarr, 2005; Loomis, Klatzky, Golledge, & Philbeck, 1999).

All else being equal, Navon (1977) proposed that visual processing initially emphasizes global processing. Rarely, however, is all else equal. Many factors, including emotion and arousal, interrupt this *global precedence*. Arousal-inducing elements of a scene can narrow attention, leading to memory for only limited details (Easterbrook, 1959; Loftus, Loftus, & Messo, 1987; Siegel & Loftus, 1978). For example, after viewing an arousing bank-robbery scene, people remember fewer overall details (Loftus & Burns, 1982), but details within the arousing area receive heightened attention (e.g. Kensinger, Garoff-Eaton, & Schacter, 2006). Affectively laden information exists in most environments, for example, a cliff-edge setting off one's fear of heights or a city area associated with violence. The spatial locations of affectively laden information can be critical to navigation planning, through mechanisms of approach-avoidance behavior (Cacioppo & Berntson, 1994; Chen & Bargh, 1999). Crawford and Cacioppo (2002) used a statistical association technique to determine the extent to which people attended to affective information and incorporated it into their spatial representations. Their results suggested that, despite the incidental exposure to the affective information, people integrated it into their representation. Further, the tendency to do so appeared strong; affective information was incorporated even when the information varied on other dimensions and/or was weakly correlated.

A distinction must be made between processing emotionally arousing information and processing information when one is in an emotionally aroused state. Emotionally arousing information can capture attention and narrow focus. In contrast, inducing an emotionally aroused state can reinforce global processing. High arousal increases false memory, a phenomenon driven by global, gist processing (Corson & Verrier, 2007; Mahoney et al., 2012; Storbeck & Clore, 2005). More chronic arousal, such as that experienced by individuals with post-traumatic stress disorder with heightened basal arousal levels, also leads to global processing advantages (Vasterling,

Duke, Tomlin, Lowery, & Kaplan, 2004). Applying this to spatial cognition, Brunyé, Ditman, et al. (2009) and Brunyé, Mahoney, et al. (2009) explored how emotional valence and arousal affect spatial memory. They induced a positive or negative mood with either a high or low level of arousal using international affective picture system (IAPS) pictures (Bradley & Lang, 2006). The participants then studied an unfamiliar campus map and completed memory assessments. The results showed that high arousal, regardless of valence, led to globally focused mental representations, as seen through amplification of the symbolic distance effect. More recent work examining effects of arousal on attention support this finding (Mahoney, Brunyé, Giles, Lieberman, & Taylor, 2011; McConnell & Shore, 2011).

Arousal and emotion affect attention, which in turn cascades to other, higher-order processes, such as spatial cognition. This has clear implications for spatial learning in applied contexts. The emergency responder may navigate differently to a call depending on whether she/he is in a high stress state or not. These navigation choices may then affect how quickly she/he arrives at the call or how detours are handled. Some details may not be processed if learned in a high arousal state, affecting later way-finding. The effect of arousal/emotion on attention has implication for other situations as well (Eysenck, 1976), including eyewitness testimony (Christianson, 1992).

2.4. Technology for and During Navigation: Attention Distractors

The use of technology has increased dramatically. Simply count the number of people staring at a smart phone next time you are in a crowded train station. This technology includes devices used for and during navigation. Regardless of whether the technology is aiding navigation or not, it diverts attention from the navigation task and the environment. Individuals engaged in a telephone conversation while driving are four times more likely to be involved in a serious accident (Strayer & Johnston, 2001). While the more dire consequences of this diversion (e.g. the 2009 Boston trolley accident caused by the driver texting; Valencia & Bierman, 2009) receive prominent news attention, consequences of this technology for spatial cognition have only recently received research attention.

Cell phone use induces inattention blindness to information in the environment (Hyman et al., 2010; Strayer & Drews, 2007). Attention gets diverted from the environment to information imparted by the device. Strayer and Drews (2007) showed the memorial consequences of this inattention blindness. Using eye-tracking, they noted which environment features participants

in a driving simulator fixated. Participants showed reduced recognition for objects their eyes had rested on when simultaneously engaged in a cell phone conversation than when just driving. In the fourth study of this work, they found that 33% more drivers missed their designated highway exit when on the cell phone than when they were not. This inattention blindness is not limited to driving where vehicle speed may limit how long environment information is available to process and driving conditions may increase overall cognitive load. Hyman et al. (2010) examined attentional consequences of cell phone use while walking. They tracked students crossing a college plaza who were or were not on a cell phone. Within the plaza, they set up an unusual event, a unicycling clown. Only 25% of participants on cell phones noticed the clown, compared to 51%, 61%, and 71% of people walking alone, listening to a music device, or walking with a friend, respectively. Their findings are consistent with other evidence that divided attention disrupts walking, the consequences of which may make people miss seeing novel stimuli or cross streets unsafely (Bungum, Day, & Henry, 2005; Hatfield & Murphy, 2007; Nasar, Hecht, & Werner, 2008).

Having seen a popularity surge in 2000, global positioning system (GPS) use continues to grow (James, 2009). Unlike talking or texting on one's cell phone, GPS systems have a navigational purpose. However, if one's long-term goal involves easily and automatically navigating through one's home environment, outcomes of using an in-vehicle navigation system appear antithetical to this goal. Memory for an environment learned while using a navigational aid is deficient compared to that developed without an aid (Aporta & Higgs, 2005; Burnett & Lee, 2005; Gardony, Brunyé, Mahoney, & Taylor, 2012b). Attention divided between the navigation aid and the environment contributes to this memory deficit (Fenech, Drews, & Bakdash, 2010; Gardony et al., 2012b). Divided attention disrupts the acquisition of spatial knowledge at both the route and survey levels (Albert, Reinitz, Beusmans, & Gopal, 1999).

Despite the memory deficits, navigational aid use continues to increase, perhaps driven by a perception that one will reach their destination with greater certainty and efficiency. However, recent work suggests that the navigational help such aids is limited. Gardony et al. (2012b) had participants navigate between 10 pairs of predefined locations in virtual environments, either aided or unaided. Prior to navigating, all participants studied a map for one minute. Only for navigation between the first pair of locations did aided navigation lead to more efficient navigation. As soon as the second pair, unaided navigation was as efficient as aided; in other words, unaided

navigators quickly develop the requisite spatial knowledge to guide them to goal locations. Work on GPS use and memory is one example of effects of cognitive off-loading onto technology, but one with high ecological validity that can contribute to a general understanding of cognitive off-loading.

In-vehicle technology, whether related or not to driving and navigation, divides attention with consequences ranging from minor to disastrous. For spatial cognition, dividing attention manifests in a reduced ability to bind location and identity information, thus leading to spatial memory deficits. Current studies in our laboratory focus specifically on divided attention and its effects on spatial memory. These studies, like others examining the role of attention in other, often more complex cognitive tasks, use dual-task paradigms, wherein participants engage in multiple tasks simultaneously. The design of these tasks and the purported cognitive processes they engage additionally reveal details of these cognitive processes (Baddeley, 1992).

2.5. Conclusions: Attention and Spatial Cognition

Without attention, little information undergoes further processing (e.g. Lachter et al., 2004). Spatial cognition research, some of which has been reviewed in this chapter, has shown how changes in attention affect the information incorporated for further use. One benefit of exploring attention within the context of spatial cognition comes from the ability to compare someone's mental representation to aspects of the physical spatial information being represented. Because of the concrete nature of the physical information, this comparison is less ambiguous than situations where multiple interpretations or inferences might come to bear.



3. MEMORY AND SPATIAL COGNITION

Memory research covers a broad range of specific issues, many which have captured prominent research attention. It is probably the cognitive sub-field in which the most research has been devoted to developing theoretical positions and computational models. However, the majority of these models rely on data from verbal learning paradigms. The generalizability of these models would be enhanced by considering results from spatial memory research.

3.1. Working Memory Involved in Spatial Cognition

Of the memory processes subserving cognition in general, working memory has received abundant research attention. Baddeley's working memory model (Baddeley, 1992, 2002) divides working memory into specialized subsystems.

The most recent conceptualization (Baddeley, 2002) defines four subcomponents, including the articulatory loop, the visuospatial sketchpad, the executive, and the more recently introduced episodic buffer. Dual-task studies have assessed the role of these subcomponents in higher-order cognitive tasks, including reading comprehension (e.g. Kintsch, 1994; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000), problem solving (e.g. Hambrick & Engle, 2003), reasoning (e.g. Gilhooly, 2004), and procedure learning (Brunyé, Taylor, Rapp, & Spiro, 2006). These studies show the complex interaction of these subcomponents. For example, we explored working memory processes in learning procedures from single and multimedia sources (Brunyé et al., 2006). Participants learned how to assemble small toys while undertaking one of the four secondary tasks. These tasks involved the articulatory loop, the visuospatial sketchpad, or central executive resources involving random (verbal or spatial). The results showed articulatory loop processing with text instructions and a visuospatial sketchpad role in picture processing. Interestingly, the central executive secondary tasks, whether verbal or spatial based, interfered with multimedia processing. The domain-specific aspects of the central-executive tasks did not selectively interfere with text or picture processing in isolation. These results suggest a choreographed interaction of working memory in higher-order cognition. Understanding how these roles shift in a variety of tasks may lead to a greater understanding of working memory as a whole.

In spatial cognition, the division of responsibility between the subcomponents of Baddeley's (1992, 2002) working memory model is not completely intuitive. The important role of the visuospatial sketchpad is both expected and clear. Spatial cognition employs visuospatial working memory (e.g. Garden, Cornoldi, & Logie, 2002). However, the role of the articulatory loop appears to be greater than one might predict. As discussed earlier, spatial information involves identity and location. People use verbal information to name landmarks, but the role of verbal processing goes beyond this. Less intuitive is the role verbal processing plays in route learning. Work examining route learning from descriptions (Brunyé & Taylor, 2008; De Beni, Pazzaglia, Gyselinck, & Meneghetti, 2005), navigation (Garden et al., 2002), or virtual navigation (Meilinger, Knauff, & Bulthoff, 2008) all implicate verbal working memory processes. Route learning, like language processing, involves serial integration. Sequence monitoring requires continuous updating of sequential representations within working memory. While this updating is a proposed central executive function (e.g. Brunyé et al., 2006; Miyake & Shah, 1999), the articulatory loop may also be involved. Evidence for this supposition comes from work showing

that verbal representation maintenance is key to effective use of cognitive resources during task execution (Pearson, Logie, & Gilhoolly, 1999). Also supporting this view are findings showing that the role of verbal working memory in route learning appears particularly important for individuals with low spatial ability (Garden et al., 2002). Finding a strong role for verbal working memory in spatial cognition extends understanding of both spatial cognition and verbal working memory.

Work exploring the role of the central executive in spatial cognition makes a similar knowledge extension. In the research literature as a whole, less is known about the central executive, but is thought to “oversee” the other working memory subsystems, integrating information from them (Baddeley, 1996; Baddeley, Emslie, Kolodny, & Duncan, 1998; Duff, 2000; Miyake & Shah, 1999). One type of integration updating serially changing information appears to fall under the central executive’s purview. Brunyé and Taylor (2008) found that central executive mechanisms played a role in developing spatial mental models, particularly from route information, and also in retrieving route-based information.

In summary, spatial cognition not only involves spatial processing but it also uses the processing power of other working memory subcomponents (Baddeley, 1992). The choreographed role of working memory changes with domains and with tasks within domains. Understanding how the working memory subcomponent roles change by domain and task provides a better understanding of working memory. Spatial cognition work has added to this understanding. Verbal working memory processes (articulatory loop) come into play to a greater extent when learning or remembering routes, most likely related to the serial nature of routes. Central executive mechanisms related to serial updating and integration, play a key role when developing spatial mental models (Brunyé & Taylor, 2008). The range of working memory resources garnered for spatial cognition tasks suggests the following: Work exploring the relative role of the working memory subcomponents in different spatial cognitive tasks may yield bidirectional understanding. In other words, such work may further explicate the role of working memory in spatial cognition and may further explicate general working memory mechanisms and their interactions.

3.2. Structuring Spatial Memory

With the amount of information people hold in memory, even if only for a short duration, it may not be surprising that organizing the information yields increased memory output. In an early example showing spontaneous

memory organization, [Bousfield \(1953\)](#) had participants study word lists containing semantic associates. The associates appeared randomly distributed in the learning list. Recall order, however, showed that participants clustered words into associates, rather than maintaining learning order. A huge body of subsequent research has specifically explored semantic memory organization resulting in many suggestions for associative network models that vary in their specifics.

Spatial memory involves associations. People associate locations when they experience them closely, in space or time ([Curiel & Radvansky, 1998](#); [Naylor-Emlen & Taylor, 2009](#)). Reinforcing these spatial and temporal associations leads them to be grouped into categories, that is, repeated associations in varied contexts. People use roads ([McNamara et al., 1984](#)) and even artificial boundaries ([McNamara, 1986](#)) as dividing lines between categories. However, packaging associations in categories can lead to memory errors and distortions. Categorically grouped locations are perceived as more similar to one another and more distinct from locations in different categories ([Hirtle & Jonides, 1985](#); [Maki, 1982](#); [McNamara, 1986](#); [McNamara et al., 1989](#)). Behaviorally, this means people misremember relative location and distance, biasing it toward central tendencies of the category ([Stevens & Coupe, 1978](#)). Categorical memory effects like these are not limited to spatial memory and have been shown in a variety of contexts. For example, social cognition shows parallel findings related to stereotyping. People categorize others based on various features, including name, skin color, hair color, clothing, and/or mannerisms. These features may activate race, ethnicity, gender, and/or age categories and lead to category-based errors, such as failure to differentiate individuals within a social category (e.g. [Hamilton, 1981](#); [Taylor, 1981](#)). Category-based memory errors also form the basis of a whole false memory literature based on the DRM paradigm (e.g. [Roediger & McDermott, 1995](#)). With the DRM paradigm, people study word lists within which all items are associated to another, unrepresented word. In integrating the associations, people then falsely report having studied the unrepresented word.

Spatial memory organization can also be examined in combination with other, nonspatial categories. Nonspatial features of an environment can be incorporated into one's cognitive map along with spatial features ([McNamara, Halpin, & Hardy, 1992](#); [McNamara & LeSueur, 1989](#)). For example, people cluster locations based on their function ([Hirtle & Jonides, 1985](#); [Merrill & Baird, 1987](#)) or physical similarity ([Hirtle & Kallman, 1988](#)), and the semantic category into which they fall ([Hirtle & Mascolo, 1986](#)).

We examined how spatial and social categories might interact (Maddox et al., 2008). This question is particularly interesting because the connection between a location and a social category is indirect; it is mediated through a person associated with the category. This work showed that racial associations to locations led to spatial and social category effects. Participants confused individuals of the same race as being associated with a particular location. Further, two locations associated with individuals of the same race were perceived as closer together than locations associated with different races. Spatial memory may be unique in affording the opportunity to examine how these mediated categories interact. Looking across memory research, categorical memory organization appears to be domain general. Yet, research on categorical organization within each domain can yield a more complete understanding of memory organization.

Memory can be organized using other types of frameworks, as well. For example, event memory often evokes event schemas or scripts (e.g. Bower, Black, & Turner, 1979; Zacks & Tversky, 2001). These schemas derive from generalization across instances of an event type. For example, we have a script for what happens during a restaurant visit and evoke that script when understanding new restaurant events (Bower et al., 1979). Generalization across spatial situations can also set up a framework for understanding and remembering spatial layouts. Recent work in our laboratory suggests that people preferentially build cognitive maps with a north-up or north-forward orientation (Brunyé et al., 2012; Brunyé, Gagnon, et al., 2012; Brunyé, Walters, et al., 2012; Brunyé, Mahoney, Gardony, & Taylor, 2010). This framework derives from experiences with maps, which are most commonly oriented with north-up (forward). This framework, then, appears to structure spatial memory developed through other spatial learning situations. Notably for our studies, participants learned environments either through spatial descriptions or virtual navigation, but not from maps. Our critical manipulation was the direction one entered the environment, either as described (spatial descriptions) or as designated with a compass (virtual navigation). All other aspects of the environments were identical, save for this manipulation. Results suggested faster environment learning and improved memory when people entered the environment aligned with a north-up framework. This suggests a schema-like framework, with a specific orientation, onto which specific locations can be organized.

To be effective, information needs to be organized in memory. The ways people organize spatial memory is not unique to spatial memory; they have parallels to how other information is organized in memory. Specifically,

spatial cognition research shows organization based on categories and on other schema-like structures, much like memory for events, discourse, and even word lists.

3.3. Conclusions: Memory and Spatial Cognition

The long, illustrious history of memory research has provided important insights into one of human-kind's most important cognitive functions—memory. Much of this work has involved memory for verbal material or pictures. More recently, bringing research from other domains, such as spatial cognition, has begun to contribute to a more general understanding of memory processes. This relationship is bidirectional as understanding of spatial cognition clearly benefits an understanding of basic memory processes.



4. REPRESENTING SPATIAL INFORMATION: EMBODIMENT

Theories of how humans represent information can be bifurcated into abstract/amodal views (Fodor, 1975; Pylyshyn, 1984) and more recent embodied views (Grafton, 2009; Lakoff, 1988; Wilson, 2002). Embodied views posit that we represent information through visual and motor information available when learning (e.g. Glenberg, 1997; Zwaan, 2004). Then, when recalling information, people evoke perceptual and/or action simulations. Some experimental evidence supports such simulations. When people think, read, or remember information, they activate associated sensory representations in modalities including, but not limited to vision (e.g. Horton & Rapp, 2003; Yaxley & Zwaan, 2007; Zwaan, Stanfield, & Yaxley, 2002), touch (Brunyé, Gagnon, et al., 2012; Brunyé, Walters, et al., 2012), and hearing (Brunyé, Ditman, Mahoney, Walters, & Taylor, 2010). They also activate motor simulations that can influence concurrent actions. For example, people perform actions congruent with those they are thinking about faster than incongruent actions, a phenomenon referred to as the *Action Compatibility Effect* (ACE; e.g. Glenberg & Kaschak, 2002; Tucker & Ellis, 2004; Zwaan & Taylor, 2006). These actions can also involve eye movements; readers move their eyes in directions consistent with described directions (Spivey, Tyler, Richardson, & Young, 2000). Neuroimaging data also support perceptual and motor simulations, with overlapping activation areas when people read or otherwise think about actions relative to actually performing them (e.g. Pulvermuller, 2008; Tettamanti et al., 2005). While much

of this evidence comes from research focused on understanding language comprehension, representational questions from spatial cognition can further explore issues of representational embodiment.

Spatial cognition research has only recently begun to explicitly explore embodiment. Although an explicit focus on embodiment is only recent, earlier spatial cognition work suggested embodiment. For example, classic mental imagery work by [Kosslyn, Ball, and Reiser \(1978\)](#) showed that mental images preserved metric information. The embodiment aspect of this work comes from their experimental paradigm—visual scanning. The time participants took to visually scan their mental image increased with increased distance, consistent with other studies showing embodiment of eye movements ([Spivey et al., 2000](#)). Showing evidence for visual simulation, [Kosslyn \(1975\)](#) showed an effect related to object size. Participants had more difficulty verifying a visual feature of an object when they imagined a small versus a large version of the object. For example, if asked to imagine a rabbit, the size one visualizes the rabbit changes with context. The rabbit will be smaller if imagined next to an elephant than if imagined next to a housefly. Perceptually, it is more difficult to see details on something small than on something large.

Anecdotal evidence also suggests that retrieving spatial information evokes actions associated with the representation. Think about gestures people make when giving directions or describing a spatial layout. Alternatively, try describing a spatial layout while your hands are otherwise occupied. It is difficult to describe a spatial layout without gesturing (some people even invoke head [or foot] movements if unable to use their hands). Anecdotes such as these have had empirical support. [Emmorey, Tversky, and Taylor \(2000\)](#) analyzed spontaneous gestures of individuals describing a map from memory. Important to the present discussion, all participants gestured during this task, one with her feet. While the mapping between hand movements and spatial representations might not be as direct as with other actions (e.g. grasping objects), spontaneous gesturing when describing space, even to an imaginary conversational partner, strongly suggests a connection between hand movements and spatial representations ([Kita & Özyürek, 2003](#)). Thus, early spatial cognition/imagery work and anecdotes suggest perceptual and/or motor simulations associated with spatial knowledge.

Examining spatial representations affords an opportunity to explore how the representation changes as knowledge develops. People learn new environments all the time. This question is more difficult to explore with language processing, since language learning begins at a very young age.

While the specific content of an utterance or discourse passage changes, the linguistic elements do not. We specifically explored how embodiment in spatial representations might change as knowledge develops (Wang, Taylor, & Brunyé, 2012). Theories of spatial knowledge development suggest changes in representational embodiment with time and experience. Classic work by Siegel and White (1975) proposed that spatial representations built through navigation progress systematically from having a landmark to a route to a configural base. Others debate this systematic progression, suggesting instead that these information types build in parallel (Evans, Marrero, & Butler, 1981; Hermer & Spelke, 1994, 1996; Montello, 1998; Yeap & Jefferies, 2000). If one believes that spatial representations preserve perceptuomotor information about environmental experiences (Regier & Carlson, 2001; Taylor & Brunyé, 2012; Tversky, 2005; Tversky & Hard, 2009), rather than being isomorphic maplike representations (Tolman, 1948), then the two viewpoints on spatial knowledge development make different predictions in terms of embodiment. Serial development of spatial memory (Siegel & White, 1975) would predict changes in embodiment with time and experience. Parallel development (e.g. Montello, 1998) would suggest evidence of embodiment early in development, although it may still refine with time and experience.

Proximity between locations within an environment should also differentially involve embodiment. People more likely interact perceptually and motorically with spatially proximal, compared to distant, buildings. One can look around and see the relative location of two adjacent buildings, maybe with a slight turn of the head. One can point directly at a nearby location. In these cases, people associate locations relative to coordinate axes of their body (e.g. Franklin & Tversky, 1990; Longo & Lourenco, 2007; Witt, Proffitt, & Epstein, 2004). In contrast, people cannot make direct perceptuomotor associations between distant landmarks and instead may resort to a more abstracted representation (Foo et al., 2005; Siegel & White, 1975). Route planning studies suggest that way finders use visual details for nearby navigation, but turn to a coarse imagelike representation for planning routes to more distant locations (Hölscher, Tenbrink, & Wiener, 2011; Wiener & Mallot, 2003). Together, these findings suggest that people are less likely to make associations between body motions and/or perceptions and distant locations. As such, evidence of embodiment, when thinking about distant location, may either show no or weak evidence of embodiment.

Our approach is built from ACE paradigms (Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006), driven conceptually by the link between gestures and spatial information retrieval (Emmorey et al., 2000). To access action

compatibility, we used an online measure of cognitive processing, mouse tracking, implemented through *Mouse Tracker* (Freeman & Ambady, 2010). Mouse Tracker records the mouse's real-time x-, y-coordinates at a 60- to 75-Hz sampling rate, providing sufficiently rich spatial and temporal fidelity to discover perceptuomotor traces in spatial memory retrieval. We used an online measure, because it more likely reflects first-pass, immediate task analysis (e.g. Marslen-Wilson & Tyler, 1976), whereas offline measures reflect the end product of cognitive processing. Online and offline measure comparisons suggest that they tap different aspects of cognitive processing (e.g. Kempler, Almor, Tyler, Anderson, & MacDonald, 1998).

Our study assessed that spatial knowledge developed through real-world interactions with an environment, in particular college students' knowledge of their own campus. We recruited undergraduates who varied in experience with their campus (first-year students vs. seniors) and asked them to verify relative spatial locations between buildings. Some buildings were close together and others were far apart. During the verification procedure, a participant saw a building name, then the name of a second building, and finally a spatial term (*left, right, front, back*). After the spatial term, the mouse became active. Participants had to click a "yes" button if the spatial term correctly described the second building's location relative to the first and click "no" if it did not. The mouse became active at the bottom center of the screen and the "yes" and "no" buttons appeared in the upper-left and right corners, counter-balanced across participants (see Figure 3.3 for a schematic of this paradigm).

With response buttons on the right and the left, we were primarily interested in trials using terms *left* and *right*. Critical in this design is the relationship between the direction of mouse movement to respond (to the right or left side of the screen) and the relative location between the campus buildings (to the right or left). For compatible trials, the movement direction and relative location corresponded (e.g. both to the right). For incompatible trials, the movement direction and relative location were opposite (e.g. move to a button on the right, but the location is to the left). An ACE should be evident when comparing mouse trajectories between compatible and incompatible trials. Specifically, the area under the curve (AUC), relative to the idealized straight-line response, should be greater for incompatible than compatible trials.

Our results showed embodiment in spatial information retrieval, consistent with the ACE predictions (see Figure 3.3 for predictions), differences in embodiment as a function of both proximity and development, operationalized through environment familiarity. Responses to incompatible trials

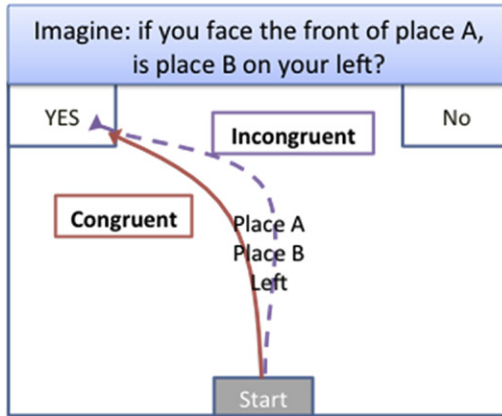


Figure 3.3 Schematic of Q. Wang et al. (2012) mouse tracking task, with hypothesized ACE results. Figure courtesy of Q. Wang. For color version of this figure, the reader is referred to the online version of this book.

veered in the direction defined by the spatial relation, leading to a greater AUC. Responses to compatible trials had a more direct trajectory. This ACE was evident for proximal, but not for distant locations, in line with predictions.

However, the embodiment effects appeared to have a different basis, depending on the extent of knowledge development. This interpretation arises from a close examination of trials presenting incorrect spatial relations. These trials allow us to distinguish between incompatibility based on the spatial relation and incompatibility based on linguistic information. For participants to verify a spatial relation, we had to present that relation to them. We did so through language, using terms *right* and *left*. Embodiment of language (Glenberg & Kaschak, 2002) predicts ACEs to these spatial terms (Coventry & Garrod, 2004). Thus, for these particular trials, ACEs based on the spatial terms and those based on the spatial representation would be at odds. Our results suggest that well-developed spatial knowledge leads to ACEs based on the spatial representation, but ACEs when spatial knowledge is less developed have a linguistic basis. High-knowledge participants showed greater AUC when the response direction conflicted with the spatial relation between the buildings; low-knowledge participants showed greater AUC when the response direction conflicted with the spatial term. Further, the real-time data output from Mouse Tracker allowed us to look at the time course of participants' responses. For high-familiarity participants, ACEs emerged later in the response; for low-familiarity participants, the ACE started early and extended throughout the response period. This suggests more immediate and

lasting action simulation for spatial language and that accessing spatial representations takes longer than accessing a spatial term's meaning. Intuitively, these different time courses make sense. People would have had a lifetime of experience with the spatial terms and less so with the Tufts campus.

To summarize, spatial representations built through interactions with an environment show evidence of embodiment, particularly when the environment is well known. Retrieving spatial knowledge of a well-learned environment appears to activate perceptual and motor simulations that reflect the sensorimotor experiences one had when learning the environment. Further, aspects of the environment that afford greater sensorimotor interaction showed greater evidence of embodiment. Specifically, landmarks close together showed ACEs, while those far apart, for which sensorimotor interaction would be limited, did not. With less experience, motor simulations arose from other, well-learned associations, in particular perceptuomotor associations to spatial terms (e.g. right). Our mouse tracking measures also showed that the ACEs arising from spatial representations and from linguistic associations unfold differently over time. These results further suggest that a spatial representation is not a unified thing, but instead made up of multiple, interrelated representations, perhaps akin to the *cognitive collage* suggested by Tversky (1993). Further, the elements and/or strength of the elements within this collage appear to change as knowledge develops. Using spatial cognition allowed this examination of how representations can change as knowledge develops.

4.1. Embodiment: Spatial Reference Frames and Representational Perspective

Our mouse tracking work not only supported embodiment in spatial representations but also suggested that not all aspects of spatial representations involve sensorimotor simulations. This also suggests that representations of nonspatial information may have both embodied and nonembodied aspects and that understanding representational issues in general involves predicting when representations will and will not involve perceptual and/or motor simulations. We further explored this question, drawing on other spatial cognitive work.

Other aspects of spatial representations that have implications for embodiment involve the reference frames and perspectives underlying and used to retrieve from representations. Spatial information can be represented using different reference frames (e.g. Carlson-Radvansky & Irwin, 1993, 1994; Carlson-Radvansky & Jiang, 1998; Levinson, 1996) and from different

perspectives (Perrig & Kintsch, 1985; e.g. Taylor et al., 1999; Taylor & Tversky, 1992; Thorndyke & Hayes-Roth, 1982). *Reference frames* provide structure for locating a specific object. Levinson (1996) outlines three spatial reference frames used in language: *intrinsic*, *relative*, and *absolute*. An intrinsic frame uses an object-centered (or other person-centered) coordinate system (e.g. the car is parked in front of the house). A relative frame uses the speaker's coordinate system (e.g. my water glass is to the right of my place, from my perspective). An absolute frame evokes an environment-centered coordinate system. When locating multiple objects, such as those comprising a real-world environment, perspectives structure how we think about and connect these locations.

“*Perspective*” reflects the viewpoint that is both taken on an environment and used to reference locations. Three primary spatial perspectives have been studied: *survey*, *route*, and *gaze*. People gain a survey perspective by studying a map or finding a vantage point above the environment (e.g. atop a hill or tree, or in an airplane). Object locations are referenced to other known locations, generally using canonical axes. In contrast, a *route perspective* reflects movement through an environment. Object locations are referenced to the current location of an imagined observer along an imaginary tour of an environment (Perrig & Kintsch, 1985; Taylor & Tversky, 1992; Thorndyke & Hayes-Roth, 1982). The route perspective is congruent with the direct experience one has navigating in an environment. The third, *gaze perspective* arose from research showing that imaginary tours can involve eye movement only, when locations are all visible from a single viewpoint (Ehrich & Koster, 1983; Levelt, 1982, 1989). Gaze tours, like a survey perspective, relate locations to other locations, but like a route perspective, do so from a within-environment vantage point. The following section considers how varied reference frames might predict the degree to which people embody perceptuomotor information in spatial representations.

4.1.1. Embodiment and Use of Alternative Reference Frames

Spatial reference frames and perspectives have implications for embodiment. Different reference frames give different visual vantage points on an environment. These reference frames can be evoked through the pronouns used in narratives. Brunyé, Ditman, Mahoney, Augustyn, and Taylor (2009) examined whether different pronouns evoked different visual representations of an action. We compared pronoun-defined intrinsic and relative reference frames. Participants verified whether pictures accurately reflected described actions (e.g. You are peeling the cucumber). Critically, the picture either showed the action or did not and either matched the spatial perspective

evoked by the pronoun (*you* vs. *he*) or did not (Figure 3.4). Results suggest that pronouns evoke different reference frames; participants responded faster to pictures matching the frame implied by the pronoun. Follow-up work further suggested embodiment tied to the reference frame (Ditman, Brunyé, Mahoney, & Taylor, 2010). This work examined whether readers spontaneously simulate described actions and whether this is mediated by the reference frame evoked. We hypothesized that embodiment would more likely accompany actions described from one's own reference frame. In other words, participants should more likely simulate actions preceded by the pronoun *you*. Participants again read actions described using different pronouns (I, you, he). After a delay, participants verified the action or another scenario detail. Results showed better memory for actions preceded by *you*. This pronoun-linked memory difference was limited to the action and was not in evidence for other scenario details. Important in these finds is the fact that the study used a memory task that did not promote mental imagery. Similarly, Kessler and Thomson (2010) found evidence of spontaneous movement simulation when taking another's reference frame. Thus, embodied simulations seem to arise more so from one's own perspective.

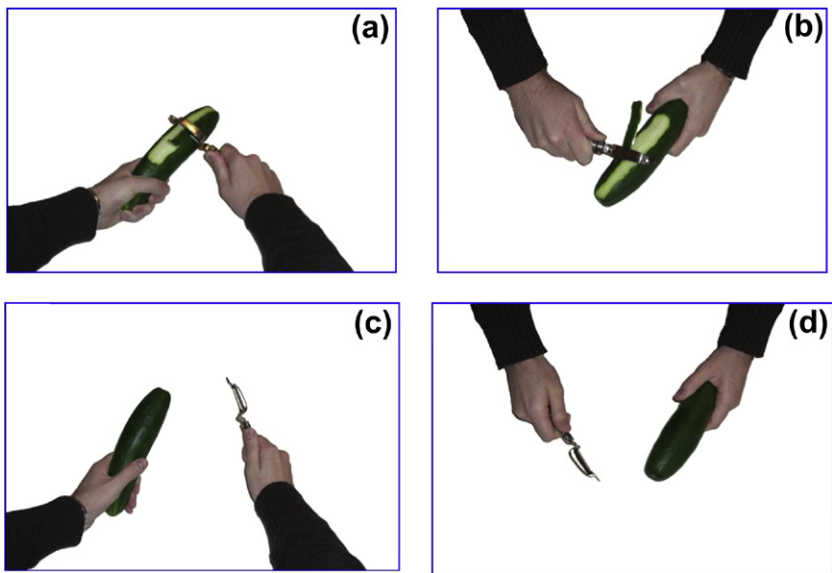


Figure 3.4 Example pictures from the study of Brunyé, Ditman, et al. (2009) and Brunyé, Mahoney, et al. (2009) either showing an action (a and b) or not (c and d) and either consistent with the pronoun “you” (a and c) or the pronoun “he” or “she” (b and d). For color version of this figure, the reader is referred to the online version of this book.

Other work has examined how embodiment of action changes spatial reference frame use. Tversky and Hard (2009) explored how perceived action affects conceptualization of a spatial layout. Participants described an object's location relative to another based on a picture. They saw one of three pictures: one depicting only the objects, one depicting a person looking at the objects, and a third depicting a person about to act on one of the objects. Merely having a person in the picture led participants to use that person's reference frame to a greater extent. Showing evidence of embodiment, participants were even more likely to use this person's reference frame if he/she appeared about to act on the objects. In a second study, participants saw the person acting on the objects and responded to one of four questions. The questions either emphasized the person or did not and either emphasized the action or did not. Questions emphasizing the action elicited more descriptions using the other person's perspective, and action questions mentioning the person showed the greatest use of the other person's perspective. These results suggest that understanding of a spatial scene changes when in service of understanding another's actions. People represent someone else's reference frame when interpreting their actions at particular locations. These findings also have implications for other types of cognitive processing, notably both discourse understanding and interpretation of social situations.

4.1.2. Embodiment of Route and Survey Perspectives

Spatial perspectives should also differentially evoke embodiment. In particular, a route perspective, because it involves a mental tour, should show more evidence of embodiment than a survey perspective, which can more likely be represented with a static cognitive map (Tolman, 1948). Several studies from our laboratory suggest this to be the case. Some of this work involves route selection (Brunyé, Mahoney, Gardony, et al., 2010). In a series of studies, participants selected the "best" route between map locations, describing their choice to the experimenter. The critical trials involved dilemmas, wherein the two primary route choices were equidistant. On half of these critical trials (east–west dilemmas), the two locations were situated north–south of one another, and one route primarily went east and the other primarily west. For the other half of the critical trials (north–south dilemmas), the locations were situated east–west of one another, and one route primarily went north and the other primarily south (Figure 3.5). Results showed 50–50 selection for east–west dilemmas, but for north–south dilemmas participants preferred southern routes at a ratio of greater than 2 to 1.

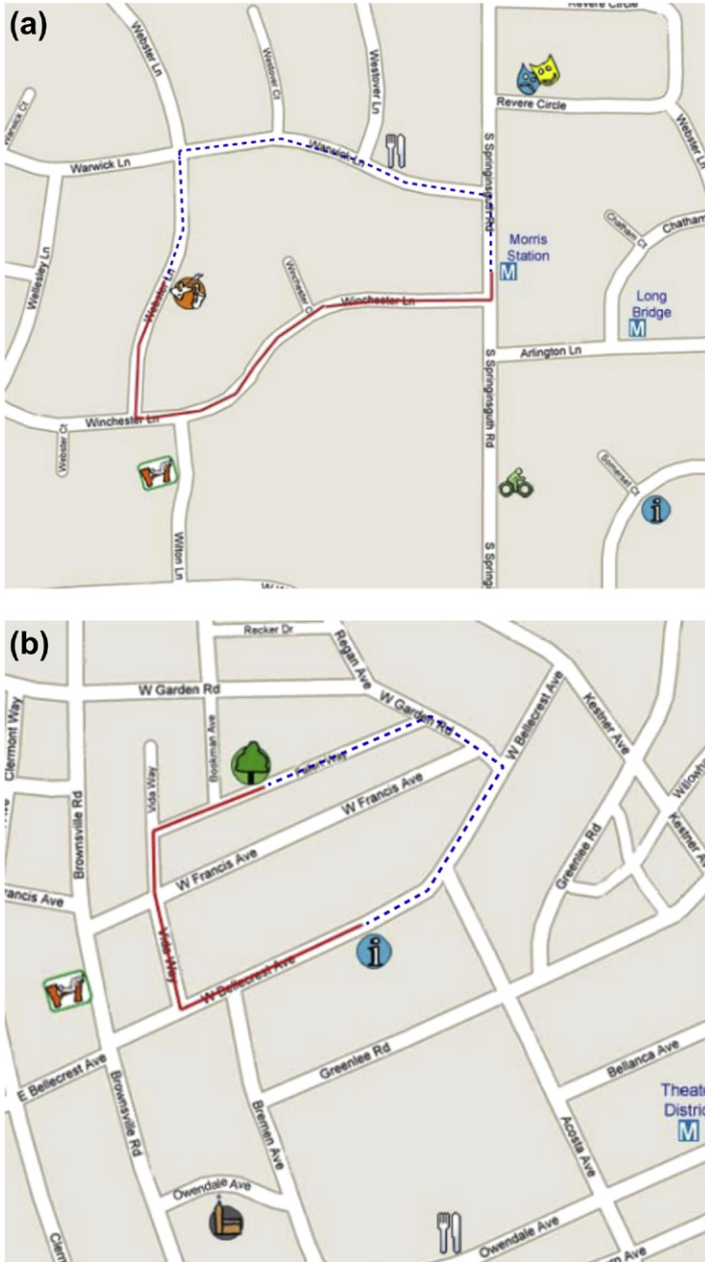


Figure 3.5 North–south (a) and east–west (b) dilemma trials from the study of Brunyé, Ditman, et al. (2010) and Brunyé, Mahoney, et al. (2010). Note that the participants only saw the landmarks and roads, but did not see the specific routes marked on the map as they are depicted here. For color version of this figure, the reader is referred to the online version of this book.

The southern route bias, however, differed as a function of spatial perspective. Notably, this southern route preference only emerged when participants described their selection from a route perspective. When describing it from a survey perspective, they had 50–50 selection of northern and southern routes. A second experiment forced a route or survey perspective and showed the southern route preference, again only when taking a route perspective. Further studies suggested that the bias arose because people perceive northern routes as going “uphill”; people rated northern, compared to southern, routes as more scenic and requiring more calories to traverse. These results suggest that different spatial perspectives evoke differential degrees of mental simulation. People prefer routes that seem less difficult to traverse. Some aspects of the study make these results particularly interesting. First, the task, selecting a route from a map, did not require mental simulation. Second, participants had no direct experience with these environments and had only seen maps of them. Yet, when describing their route choice from a route perspective, they appear to embody an impression that northern routes go uphill and will, therefore, be more difficult to travel.

Other work suggesting differential embodiment in route and survey perspectives explored how these perspectives tie to physical actions, in particular walking. A route perspective mental tour involves a mode of translation, including walking. In two studies, Brunyé, Mahoney, and Taylor (2010) assessed the relative degree of mental simulation when reading route versus survey descriptions. They did so using action-related versus action-unrelated sounds during learning. Participants read a route or survey description while listening to either footsteps on gravel or metronome sounds. The speed of the sound was either fast (3 pulses/second) or slow (1 pulse/second). For the footsteps, the two speeds implied running or walking. The speed difference had important implications for embodiment. One should be able to cover more distance when running than when walking. Dependent measures examined both encoding (reading time) and memory (statement verification for experiment 1 and map drawing and scale estimation for experiment 2). For encoding, the type of sound interacted with the spatial perspective. With metronome (non-action-based) sounds, readers modulated their reading speed based on the sound speed, faster reading with faster sounds. In contrast, footstep (action-based) sounds only affected route description reading. Participants read route descriptions faster if the sound implied running, but they read survey descriptions at the same speed whether the sounds implied walking or running.

These simulation effects carried over to memory. In particular, participants showing greater simulation, as assessed by greater influence of the action-based sound, appeared to root their mental representation in the route perspective. They had difficulty solving inferences from a survey perspective. Experiment 2 more directly assessed the implication of simulating running versus walking, by examining representation of distance. In this study, learning followed the experiment 1 procedures, but then participants drew maps and estimated the scale distance of their map. The speed of the action-based sound influenced perceptions of distance traveled. Participants drew locations further apart and gave greater scale estimates after listening to running versus walking sounds. In other words, if the sounds they heard while reading a route description simulated running, they perceived locations within the environment as further apart and the environment as encompassing a greater area. This effect was not evident for survey descriptions. These results, like the route selection results, are notable in how participants learned the environment. In both cases, participants had no physical experience within the environment. In the route selection studies (Brunyé, Mahoney, et al., 2010), participants only saw maps. For the action-based sound studies (Brunyé, Mahoney, et al., 2010), participants read environment descriptions. Taken together, these studies suggest that embodiment effects in spatial cognition need not evolve from direct actions and perceptions. Instead, the findings suggest that people can simulate the actions and perceptions one might have when traveling within a novel environment, most likely based on other navigational experiences.

4.2. Embodied Spatial Representations: Conclusions

Mounting evidence suggests that spatial representations are multifaceted. With respect to embodiment, research suggests that some, but not all, aspects of spatial representations evoke associated perceptual and/or motor simulations. Thinking about locations distant from one another does not bring about the embodied processing that thinking about nearby locations does. Indeed one cannot easily interact on a perceptual and motor level with locations distant from one another. Taking a mental tour of an environment (route perspective) to remember locations involves more perceptuomotor simulation than remembering locations from a bird's-eye viewpoint (survey perspective). This appears to be true whether the spatial information was learned from a route or a survey perspective. This view is in line with other proposals, drawing on spatial cognition and cognitive neuroscience findings, for multiple systems of spatial memory (e.g. Avraamides & Kelly, 2008). A multiple systems account allows the flexibility to consider the variety of

situations in which people learn about and use spatial information (Mallot & Basten, 2009). This work has implications for cognitive processing in other multifaceted situations.



5. THE COGNITION OF SPATIAL COGNITION

In this chapter, we have discussed spatial cognition research relative to three cognitive processes: attention, memory, and representation. Few would argue that understanding domain-general aspects of these processes leads to a stronger theoretical understanding of spatial cognition. To this end, bringing together, in one place, a discussion of how attention, memory, and representation issues bear out in spatial cognition has utility. We argue that discussing the reverse relationship may have additional utility. Specifically, understanding how attention, memory, and representation play out in spatial cognition can lead to a better domain-general understanding of these processes. All cognitive research has domain-specific elements, embedded in the concrete tasks participants undertake. For a domain-general model or theory of a cognitive process to be accurate, it must hold across a variety of domains. The spatial domain has not played a central role in studies of domain-general processing, but has the potential to contribute substantively.

REFERENCES

- Albert, W. S., Reinitz, M. T., Beusmans, J. M., & Gopal, S. (1999). The role of attention in spatial learning during simulated route navigation. *Environment and Planning A*, *31*, 1459–1472.
- Aporta, C., & Higgs, E. (2005). Satellite culture: global positioning systems, inuit wayfinding, and the need for a new account of technology. *Current Anthropology*, *46*(5), 729–753.
- Avraamides, M. N., & Kelly, J. W. (2008). Multiple systems of spatial memory and action. *Cognitive Processes*, *9*, 93–106.
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.
- Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559.
- Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *49*, 5–28.
- Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, *7*, 85–97.
- Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology*, *51A*, 818–852.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723–767.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology*, *49*(2), 229–240.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220.

- Bradley, M. M., & Lang, P. J. (2006). The international affective picture system (IAPS) in the study of emotion and attention. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 29–46). Oxford: Oxford University Press.
- Brunyé, T. T., Andonova, E., Neneghetti, C., Noordzij, M., Pazzaglia, F., Wienemann, R., et al. (2012). Planning routes around the world: international evidence for southern route preferences. *Journal of Environmental Psychology, 32*, 297–304.
- Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). When you and I share perspectives: pronouns modulate perspective-taking during narrative comprehension. *Psychological Science, 20*, 27–32.
- Brunyé, T. T., Ditman, T., Mahoney, C. R., Walters, E. K., & Taylor, H. A. (2010). You heard it here first: readers mentally simulate described sounds. *Acta Psychologica, 135*(2), 209–215.
- Brunyé, T. T., Gagnon, S. A., Waller, D. A., Hodgson, E., Tower-Richardi, S., & Taylor, H. A. (2012). Up north and down south: implicit associations between topography and cardinal direction. *Quarterly Journal of Experimental Psychology, 65*(10), 1880–1894.
- Brunyé, T. T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). Emotional state and local vs. global spatial memory. *Acta Psychologica, 130*, 138–146.
- Brunyé, T. T., Mahoney, C. R., Gardony, A., & Taylor, H. A. (2010). North is up(hill): route planning heuristics in real-world environments. *Memory & Cognition, 38*(6), 700–712.
- Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2010). Moving through imagined space: mentally simulating locomotion during spatial description reading. *Acta Psychologica, 134*(1), 110–124.
- Brunyé, T. T., & Taylor, H. A. (2008). Working memory in developing and applying mental models from spatial descriptions. *Journal of Memory and Language, 58*(3), 701–729.
- Brunyé, T. T., & Taylor, H. A. (2009). When goals constrain: eye movements and memory for goal-oriented map study. *Applied Cognitive Psychology, 22*, 1–16.
- Brunyé, T. T., Taylor, H. A., Rapp, D. N., & Spiro, A. (2006). Learning procedures: the role of working memory in multimedia learning experiences. *Applied Cognitive Psychology, 20*(7), 917–940.
- Brunyé, T. T., Walters, E. K., Ditman, T., Gagnon, S. A., Mahoney, C. R., & Taylor, H. A. (2012). The fabric of thought: priming tactile properties during reading influences direct tactile perception. *Cognitive Science, 36*(8), 1449–1467.
- Bungum, T. J., Day, C., & Henry, L. J. (2005). The association of distraction and caution displayed by pedestrians at a lighted crosswalk. *Journal of Community Health, 30*, 269–279.
- Burnett, G. E. (2000). “Turn right at the traffic lights”: the requirements for landmarks in vehicle navigation systems. *The Journal of Navigation, 53*(3), 499–510.
- Burnett, G. E., & Lee, K. (2005). The effect of vehicle navigation systems on the formation of cognitive maps. In G. Underwood (Ed.), *Traffic and transport psychology: Theory and application* (pp. 407–418). Amsterdam: Elsevier.
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: a critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*, 401–423.
- Carlson-Radvansky, L. A., & Irwin, D. A. (1993). Frames of reference in vision and language: where is above? *Cognition, 46*, 223–244.
- Carlson-Radvansky, L. A., & Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language, 33*, 646–671.
- Carlson-Radvansky, L. A., & Jiang, Y. (1998). Inhibition accompanies reference-frame selection. *Psychological Science, 9*(5), 386–391.
- Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24*, 403–416.
- Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Personality & Social Psychology Bulletin, 25*, 215–224.

- Christianson, S.-A. (1992). Emotional stress and eyewitness memory: a critical review. *Psychological Bulletin*, 112(2), 284–309.
- Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinions in Neurobiology*, 17, 177–184.
- Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11(9), 393–399.
- Clore, G. L., Wyer, R. S., Dienes, B., Gasper, K., Gohm, C., & Isbell, L. (2001). Affective feelings as feedback: some cognitive consequences. In L. L. Martin, & G. L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 27–62). Mahwah, NJ: Erlbaum.
- Corson, Y., & Verrier, N. (2007). Emotion and false memory: valence or arousal? *Psychological Science*, 18, 208–211.
- Coventry, K. R., & Garrod, S. C. (2004). *Saying, seeing, and acting: The psychological semantics of spatial prepositions*. New York: Psychology Press.
- Crawford, L. E., & Cacioppo, J. T. (2002). Where to look for danger: integrating affective and spatial information. *Psychological Science*, 13(5), 449–453.
- Curiel, J. M., & Radvansky, G. A. (1998). Mental organization of maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(1), 202–214.
- Dalrymple, K. A., Birmingham, E., Bischof, W. F., Barton, J. J.S., & Kingstone, A. (2010). Opening a window on attention: documenting and simulating recovery from simultaneity. *Cortex*, 47(7), 787–799.
- De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology*, 17(1), 77–95.
- Ditman, T., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2010). Simulating an enactment effect: pronouns guide action simulation during narrative comprehension. *Cognition*, 115(1), 172–178.
- Duff, S. C. (2000). What's working in working memory: a role for the central executive. *Scandinavian Journal of Psychology*, 41, 9–16.
- Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66, 183–201.
- Ehrich, V., & Koster, C. (1983). Discourse organization and sentence form: the structure of room descriptions in Dutch. *Discourse Processes*, 6, 169–195.
- Ellis, N. R., Katz, E., & Williams, J. E. (1987). Developmental aspects of memory for spatial location. *Journal of Experimental Child Psychology*, 44(3), 401–412.
- Emmorey, K., Tversky, B., & Taylor, H. A. (2000). Using space to describe space: perspective in speech, sign, and gesture. *Spatial Cognition & Computation*, 2(3), 157–180.
- Eriksen, C. W. (1995). The flankers task and response competition: a useful tool for investigating a variety of cognitive problems. *Visual Cognition*, 2(2–3), 101–118.
- Evans, G. W., Marrero, D. G., & Butler, P. A. (1981). Environment learning and cognitive mapping. *Environment and Behavior*, 13, 83–104.
- Eysenck, M. W. (1976). Arousal, learning, and memory. *Psychological Bulletin*, 83(3), 389–404.
- Fenech, E. P., Drews, F. A., & Bakdash, J. Z. (2010). *The effects of acoustic turn-by-turn navigation on wayfinding*. Paper presented at the Human Factors and Ergonomics Society, San Francisco.
- Fodor, J. A. (1975). *The language of thought*. New York: Crowell.
- Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- vs. landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 195–215.
- Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119(1), 63–76.
- Franks, J. J., Billrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer-appropriate processing (TAP). *Memory & Cognition*, 28(6), 1140–1151.

- Freeman, J. B., & Ambady, N. (2010). Mouse tracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241.
- Garden, S., Cornoldi, C., & Logie, R. H. (2002). Visuo-spatial working memory in navigation. *Applied Cognitive Psychology*, *16*(1), 35–50.
- Gardony, A., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2012a). *Efficient navigation, deficient representation: exploring how navigational aids impair spatial memory*. Paper presented at the convention of the Association for Psychological Science.
- Gardony, A., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2012b). *Efficient navigation, deficient representation: exploring how navigational aids impair spatial memory*. Paper presented at the annual convention of the Association for Psychological Science.
- Gasper, K., & Clore, G. L. (2002). Attending to the big picture: mood and global versus local processing of visual information. *Psychological Science*, *13*(1), 34–40.
- Gauvain, M., & Rogoff, B. (1986). Influence of the goal on children's exploration and memory of large-scale space. *Developmental Psychology*, *22*, 72–77.
- Gilhooly, K. J. (2004). Working memory and reasoning. In R. J. Sternberg, & J. P. Leighton (Eds.), *The nature of reasoning* (pp. 49–77). New York, NY: Cambridge University Press.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1–55.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558–565.
- Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences*, *1156*, 97–117.
- Halligan, P. W., Fink, G. R., Marshall, J. C., & Vallar, G. (2003). Spatial cognition: evidence from visual neglect. *Trends in Cognitive Sciences*, *7*(3), 125–133.
- Hambrick, D. Z., & Engle, R. W. (2003). The role of working memory in problem solving. In J. E. Davidson, & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 176–206). New York, NY: Cambridge University Press.
- Hamilton, D. L. (1981). Stereotyping and inter-group behavior: some thoughts on the cognitive approach. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 333–354). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hasher, L., & Zacks, R. T. (1978). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*(3), 356–388.
- Hatfield, J., & Murphy, S. (2007). The effects of mobile phone use on pedestrian crossing behavior at signalised and unsignalised intersections. *Accident Analysis and Prevention*, *39*, 197–205.
- Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature Neuroscience*, *370*(6484), 57–59.
- Hermer, L., & Spelke, E. S. (1996). Modularity and development: the case of spatial reorientation. *Cognition*, *61*(3), 195–232.
- Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition*, *13*(3), 208–217.
- Hirtle, S. C., & Kallman, H. J. (1988). Memory for the locations of pictures: evidence for hierarchical clustering. *American Journal of Psychology*, *101*(2), 159–170.
- Hirtle, S. C., & Mascolo, M. F. (1986). Effect of semantic clustering on the memory of spatial locations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *12*(2), 182–189.
- Hölscher, C., Tenbrink, T., & Wiener, J. M. (2011). Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition*, *121*(2), 228–247.
- Hopfinger, J. B., Buonocore, M. H., & Magnun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, *3*, 284–291.
- Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, *10*, 104–110.

- Hyman, I. E., Jr., Boss, S. M., Wise, B. M., McKenzie, K. E., & Caggiano, J. M. (2010). Did you see the unicycling clown? Inattention blindness while walking and talking on a cell phone. *Applied Cognitive Psychology, 24*, 597–607.
- Hyun, J., Woodman, G. F., & Luck, S. J. (2009). The role of attention in the binding of surface features to locations. *Visual Cognition, 17*(1–2), 10–24.
- James, W. (1890). *Principles of psychology*. New York, NY: Holt.
- James, R. (2009). A brief history of GPS. *Time, 18*.
- Johnston, J. C., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 843–856.
- Kempler, D., Almor, A., Tyler, L. K., Anderson, E. S., & MacDonald, M. C. (1998). Sentence comprehension deficits in Alzheimer's disease: a comparison of off-line vs. on-line sentence processing. *Brain and Language, 64*(3), 297–316.
- Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language, 54*, 99–112.
- Kessels, R. P. C., Hobbel, D., & Postma, A. (2007). Aging, context memory, and binding: a comparison of “what, where, and when” in young and older adults. *International Journal of Neuroscience, 117*(6), 795–810.
- Kessler, K., & Thomson, L. A. (2010). Embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition, 114*, 72–88.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294–303.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16–32.
- Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology, 7*, 341–370.
- Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*(1), 47–60.
- LaBerge, D. (1995). *Attentional processing: The brain's art of mindfulness*. Cambridge, MA: Harvard University Press.
- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958): still no identification without attention. *Psychological Review, 55*, 207–234.
- Lakoff, G. (1988). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.
- Levelt, W. J. M. (1982). Cognitive styles in the use of spatial direction terms. In R. J. Jarvella, & W. Klein (Eds.), *Speech, place, and action* (pp. 251–268). Chichester, United Kingdom: Wiley.
- Levelt, W. J. M. (1989). *Speaking: From intentions to articulation*. Cambridge, MA: The MIT Press.
- Levinson, S. C. (1996). Frames of reference and Molyneux's question: crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 109–169). Cambridge, MA: The MIT Press.
- Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology, 12*(3), 412–445.
- Light, L. L., & Zelinski, E. M. (1983). Memory for spatial information in young and old adults. *Developmental Psychology, 19*(6), 901–906.
- Loftus, E. F., & Burns, T. E. (1982). Mental shock can produce retrograde amnesia. *Memory & Cognition, 10*, 318–323.
- Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about weapon focus. *Law and Human Behavior, 11*, 55–62.

- Longo, M. R., & Lourenco, S. F. (2007). Spatial attention and the mental number line: evidence for characteristic biases and compression. *Neuropsychologia*, *45*(7), 1400–1407.
- Loomis, J. M., Klatzky, R. L., Golledge, R. G., & Philbeck, J. W. (1999). Human navigation by path integration. In R. G. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 125–151). Baltimore, MD: Johns Hopkins Press.
- Maddox, K. B., Rapp, D. N., Brion, S., & Taylor, H. A. (2008). Social influences on spatial memory. *Memory & Cognition*, *36*(3), 479–494.
- Mahoney, C. R., Brunyé, T. T., Giles, G. E., Ditman, T., Lieberman, H. R., & Taylor, H. A. (2012). Caffeine increases false memory in nonhabitual consumers. *Journal of Cognitive Psychology*, *24*(4), 420–427.
- Mahoney, C. R., Brunyé, T. T., Giles, G., Lieberman, H. R., & Taylor, H. A. (2011). Caffeine-induced physiological arousal accentuates global processing bias. *Pharmacology, Biochemistry, and Behavior*, *99*(1), 59–65.
- Maki, R. H. (1982). Why do categorization effects occur in comparative judgment tasks? *Memory & Cognition*, *10*(3), 252–264.
- Mallot, H. A., & Basten, K. (2009). Embodied spatial cognition: biological and artificial systems. *Image and Vision Computing*, *27*, 1658–1670.
- Mandler, J. M., Seegmiller, D., & Day, J. (1977). On the coding of spatial information. *Memory & Cognition*, *5*(1), 10–16.
- Marslen-Wilson, W., & Tyler, L. K. (1976). Memory and levels of processing in a psycholinguistic context. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(2), 112–119.
- Maruff, P., Danckert, J., Camplin, G., & Currie, J. (1999). Behavioral goals constrain the selection of visual information. *Psychological Science*, *10*, 522–525.
- McConnell, M. M., & Shore, D. I. (2011). Upbeat and happy: arousal as an important factor in studying attention. *Cognition & Emotion*, *25*(7), 1184–1195.
- McNamara, T. P. (1986). Mental representation of spatial relations. *Cognitive Psychology*, *18*, 87–121.
- McNamara, T. P., Halpin, J. A., & Hardy, J. K. (1992). The representation and integration in memory of spatial and nonspatial information. *Memory & Cognition*, *20*, 519–532.
- McNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 211–227.
- McNamara, T. P., & LeSueur, L. L. (1989). Mental representations of spatial and nonspatial relations. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *41*(2-A), 215–233.
- McNamara, T. P., Ratcliff, R., & McKoon, G. (1984). The mental representation of knowledge acquired from maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 723–732.
- Meilinger, T., Knauff, M., & Bulthoff, H. (2008). Working memory in wayfinding—a dual task experiment in a virtual city. *Cognitive Science*, *32*(4), 755–770.
- Merrill, A. A., & Baird, J. C. (1987). Semantic and spatial factors in environmental memory. *Memory & Cognition*, *15*(2), 101–108.
- Miyake, A., & Shah, P. (1999). Toward unified theories of working memory: emerging general consensus, unresolved theoretical issues and future research directions. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 621–640). New York, NY: Cambridge University Press.
- Montello, D. R. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments. In M. J. Egenhofer, & R. G. Golledge (Eds.), *Spatial and temporal reasoning in geographic information systems* (pp. 143–154). New York, NY: Oxford University Press.
- Moray, N. (1959). Attention in dichotic listening: affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60.

- Nasar, J., Hecht, P., & Werner, R. (2008). Mobile telephones, distracted attention, and pedestrian safety. *Accident Analysis and Prevention*, *40*, 69–75.
- Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.
- Naylor-Emlen, S., & Taylor, H. A. (2009). Contributing factors to temporal and spatial associations in mental representations for maps. *Spatial Cognition & Computation*, *9*(4), 262–286.
- Park, D. C., Puglisi, J. T., & Lutz, R. (1982). Spatial memory in older adults: effects of intentionality. *Journal of Gerontology*, *37*, 330–335.
- Pearson, D. G., Logie, R. H., & Gilhooly, K. J. (1999). Verbal representations and spatial manipulation during mental synthesis. *European Journal of Cognitive Psychology*, *11*(3), 295–314.
- Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language*, *24*, 503–518.
- Pichert, J. W., & Anderson, R. C. (1977). Taking different perspectives on a story. *Journal of Educational Psychology*, *69*, 309–315.
- Postma, A., & De Haan, E. H. F. (1996). What was where? Memory for object locations. *The Quarterly Journal of Experimental Psychology*, *49*, 178–199.
- Postma, A., Kessels, R. P. C., & Van Asselen, M. (2008). How the brain remembers and forgets where things are: the neurocognition of object location memory. *Neuroscience and Biobehavioral Reviews*, *32*, 1339–1345.
- Pouliot, S., & Gagnon, S. (2005). Is egocentric space automatically encoded? *Acta Psychologica*, *118*, 193–210.
- Pulvermuller, F. (2008). Grounding language in the brain. In M. dVega, A. Graesser, & A. M. Glenberg (Eds.), *Symbols, embodiment, and meaning* (pp. 85–116). Oxford: Oxford University Press.
- Pylshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Rayner, K. (1979). Eye guidance in reading: fixation locations within words. *Perception*, *8*(1), 21–30.
- Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: an empirical and computational investigation. *Journal of Experimental Psychology: General*, *130*(2), 273–298.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.
- Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 803–814.
- Schulman, A. I. (1973). Recognition memory and the recall of spatial location. *Memory & Cognition*, *1*, 256–260.
- Seigneuric, A., Ehrlich, M., -F, Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing*, *13*(1–2), 81–103.
- Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and general theory. *Psychological Review*, *84*(2), 127–190.
- Siegel, J. M., & Loftus, G. F. (1978). Impact of anxiety and life stress upon eyewitness testimony. *Bulletin of the Psychonomic Society*, *12*, 479–480.
- Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10, pp. 9–55). New York, NY: Academic Press.
- Simons, D. J. (1996). In sight, out of mind: when object representations fail. *Psychological Science*, *7*(5), 301–305.
- Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *7*(1/2/3), 1–15.

- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattention for dynamic events. *Perception*, 28, 1059–1074.
- Smith, E. E., & Jonides, J. (1997). Working memory: a view from neuroimaging. *Cognitive Psychology*, 33, 5–42.
- Spelke, E. S., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition*, 4(3), 215–230.
- Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: a meta-analysis. *Psychology & Aging*, 10(4), 527–539.
- Spivey, M. J., Tyler, M., Richardson, D. C., & Young, E. (2000). *Eye movements during comprehension of spoken scene descriptions*. Paper presented at the 22nd annual meeting of the Cognitive Science Society.
- Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, 10, 422–437.
- Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy: with happiness, false memory. *Psychological Science*, 16, 785–791.
- Strayer, D. L., & Drews, F. A. (2007). Cell-phone induced driver distraction. *Current Directions in Psychological Science*, 16, 128–131.
- Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, 12, 462–466.
- Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83–114). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Taylor, H. A., & Bruny , T. T. (2012). I go right, north, and over: processing spatial language. In D. A. Waller & L. Nadel (Eds.), *Handbook of spatial cognition*, (pp. 229–248). Washington, DC: American Psychological Association.
- Taylor, H. A., Naylor, S. J., & Chechile, N. A. (1999). Goal-specific influences on the representation of spatial perspective. *Memory & Cognition*, 27(2), 309–319.
- Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, 31, 261–292.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17, 273–281.
- Thomas, A. K., Bonura, B. M., & Taylor, H. A. (2012). The influence of semantic relationships on older adult map memory. *Psychology & Aging*, 27(3), 657–665.
- Thomas, A. K., Bonura, B. M., Taylor, H. A., & Bruny , T. T. (2012). Metacognitive monitoring in visuo-spatial working memory. *Psychology & Aging*, Advance online publication. doi: 10.1037/a0028556.
- Thomas, A. K., & Bulevich, J. B. (2006). Effective cue utilization reduces memory errors in older adults. *Psychology & Aging*, 21(2), 379–389.
- Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14, 560–589.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55, 189–208.
- Treisman, A. (1996). The binding problem. *Current Opinions in Neurobiology*, 6, 171–178.
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, 116, 185–203.
- Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In U. Frank, & I. Campari (Eds.), *COSIT'93: Lecture notes in computer science* (Vol. 716, pp. 14–24). Berlin, Germany: Springer.
- Tversky, B. (2005). How to get around by mind and body: spatial thought, spatial action. In A. Zilhao (Ed.), *Cognition, evolution, and rationality: A cognitive science for the XXIst century* (pp. 135–147). London, United Kingdom: Routledge.

- Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective taking. *Cognition*, 110(1), 124–129.
- Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinions in Neurobiology*, 4, 157–165.
- Valencia, M. J., & Bierman, N. (2009). *MBTA: Conductor in Boston trolley was texting his girlfriend*. : Boston Globe.
- van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29(8), 1081–1087.
- Vasterling, J. J., Duke, L. M., Tomlin, H., Lowery, N., & Kaplan, E. (2004). Global–local visual processing in posttraumatic stress disorder. *Journal of the International Neuropsychological Society*, 10, 709–718.
- Von Wright, J. M., Gebhard, P., & Karttunen, M. (1975). A developmental study of the recall of spatial location. *Journal of Experimental Child Psychology*, 20, 181–190.
- Walmsley, D. J., & Jenkins, J. (1991). Mental maps, locus of control, and activity: a study of business tourists in Coff's Harbour. *Journal of Tourism Studies*, 2(2), 36–42.
- Wang, R. F., & Simons, D. J. (1998). Perceiving real-world viewpoint changes. *Psychological Science*, 9(4), 315–320.
- Wang, Q., Taylor, H. A., & Brunyé, T. T. (2012). When going the right way is the hard thing to do: distinct phases of action compatibility in spatial knowledge development. *Acta Psychologica*, 139(3), 449–457.
- Wiener, J. M., & Mallot, H. A. (2003). 'Fine-to-coarse' route planning and navigation in regionalized environments. *Spatial Cognition & Computation*, 3(4), 331–358.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.
- Witt, J. K., Proffitt, D. R., & Epstein, W. (2004). Perceiving distance: a role of effort and intent. *Perception*, 33(5), 577–590.
- Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, 105(1).
- Yeap, W. K., & Jefferies, M. E. (2000). On early cognitive mapping. *Spatial Cognition & Computation*, 2(2), 85–116.
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127(1), 3–21.
- Zwaan, R. A. (2004). The immersed experienter: toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 44, pp. 35–62). San Diego, CA: Academic Press.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168–171.
- Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General*, 135, 1–11.



Perceptual Learning, Cognition, and Expertise

Philip J. Kellman^{*1}, Christine M. Massey^{**}

^{*}Department of Psychology, University of California, Los Angeles, CA, USA

^{**}Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA, USA

¹Corresponding author: E-mail: Kellman@cognet.ucla.edu

Contents

1. Introduction	118
2. Perceptual Learning Effects	122
3. Perceptual Learning in Mathematics: An Example	124
4. Perception, Cognition, and Learning	130
4.1. The Classical View of Perception	131
4.2. Perceptual Symbol Systems	132
4.3. Problems for Understanding Perceptual Learning	133
4.4. The Amodal, Abstract Character of Perception	134
4.5. The Selective Character of Perception	139
4.6. Common Amodal Perceptual Representations for Thought, Action, and Learning	140
4.6.1. <i>Embodied Cognition</i>	142
4.7. Implications for Perceptual Learning	143
5. Perceptual Learning and Instruction	144
5.1. Natural Kind Learning	145
5.2. Relations among Types of Learning: Toward a “Fundamental Theorem of Learning”	147
5.3. Perceptual Learning Technology	149
5.4. Elements of Perceptual Learning in Instruction	154
6. Conclusion	159

Abstract

Recent research indicates that perceptual learning (PL)—experience-induced changes in the way perceivers extract information—plays a larger role in complex cognitive tasks, including abstract and symbolic domains, than has been understood in theory or implemented in instruction. Here, we describe the involvement of PL in complex cognitive tasks and why these connections, along with contemporary experimental and neuroscientific research in perception, challenge widely held accounts of the relationships among perception, cognition, and learning. We outline three revisions to common assumptions about these relations: 1) Perceptual mechanisms provide complex and *abstract* descriptions of reality; 2) Perceptual representations are often

amodal, not limited to modality-specific sensory features; and 3) Perception is *selective*. These three properties enable relations between perception and cognition that are both synergistic and dynamic, and they make possible PL processes that adapt information extraction to optimize task performance. While PL is pervasive in natural learning and in expertise, it has largely been neglected in formal instruction. We describe an emerging PL technology that has already produced dramatic learning gains in a variety of academic and professional learning contexts, including mathematics, science, aviation, and medical learning.



1. INTRODUCTION

On a good day, the best human chess grandmaster can defeat the world's best chess-playing computer. Not every time, but sometimes. The computer program is relentless; every second, it examines upward of 200 million possible moves. Its makers incorporate sophisticated methods for evaluating positions, and they implement strategies gotten from grandmaster consultants. Arrayed against these formidable techniques, it is surprising that any human can compete at all.

If, like the computer, humans played chess by searching through possible moves, pitting human versus computer would be pointless. Estimates of human search in chess suggest that even the best players examine on the order of four possible move sequences, each about four plies deep (where a ply is a pair of turns by the two sides). That estimate is *per turn*, not per second, and a single turn may take many seconds. If the computer were limited to 10 s of search per turn, its advantage over the human would be about 1,999,999,984 moves searched per turn.

Given this disparity, how can the human even compete? The accomplishment suggests information-processing abilities of remarkable power but mysterious nature. Whatever the human is doing, it is, at its best, roughly equivalent to 2 billion moves per second of raw search. It would not be overstating to describe such abilities as “magical.”

We have not yet said what abilities make this possible, but before doing so, we add another observation. Biological systems often display remarkable structures and capacities that have emerged as evolutionary adaptations to serve particular functions. Compared to flying machines that humans have invented, the capabilities of a dragonfly, hummingbird, or mosquito are astonishing. Yet, unlike anatomical and physiological adaptations for movement, the information-processing capabilities we are considering are all the more remarkable because it is unlikely that they evolved for one particular task. We did not evolve to play chess. What explains human attainments

in chess are highly general abilities that contribute to learned expertise in many domains. Such abilities may have evolved for ecologically important tasks, but they have such power and generality that humans can become remarkably good in almost any domain involving complex structure.

What abilities are these? They are abilities of perceptual learning (PL). The effects we are describing arise from experience-induced changes in the way perceivers pick up information. With practice in any domain, humans become attuned to the relevant features and structural relations that define important classifications, and over time, we come to extract these with increasing selectivity and fluency.

The existence of PL and its pervasive role in learning and expertise say something deeply important about the way human intelligence works. What it says violates common conceptions that view perception and learning as separate and nonoverlapping processes. It is common to think of perception as delivering basic information in a relatively unchanging way. According to this view, high-level learning happens elsewhere—in committing facts to memory, acquiring procedures, or generating more complex or abstract products from raw perceptual inputs by means of reasoning processes. Contemporary experimental and neuroscientific research in perception, as well as new discoveries in PL, require revision of these assumptions in at least three ways: 1) perceptual mechanisms provide complex and abstract descriptions of reality, overlapping and interacting deeply with what have traditionally been considered “higher” cognitive functions; 2) the representations generated by these perceptual mechanisms are not limited to low-level sensory features bound to separate sensory modalities; and 3) what perception delivers is not fixed, but progressively changing and adaptive.

We return to the first two ideas later on, but consider now what is implied by the third idea, the idea of PL. We can understand the adaptive nature of our perceptual abilities by way of contrast. Suppose we developed a set of algorithms in a computer vision system to recognize faces. The system would be structured to take input through a camera and perform certain computations on that input. If it worked properly, when we used the system for the thousandth time, it would carry out these computations in the same way as it did its first time. It is natural to think of a perceiving system as set up to acquire certain inputs and perform certain computations, ultimately delivering certain outputs.

Our brains do not work this way. If recognizing faces is the task, the brain will leverage ongoing experience to discover which features and patterns make a difference for important face classifications. Over time, the

system will become selectively attuned to extract this information and take it in in bigger chunks. (This is true even for perceptual abilities which, like face perception, likely have innate foundations.) With appropriate practice, this information extraction will become faster and more automatic. The automatization of basic information pickup paves the way for the discovery of even more complex relations and finer detail, which in turn becomes progressively easier to process (Bryan & Harter, 1899). This cyclic process can be a positive feedback loop: Improvements in information extraction lead to even more improvements in information extraction. The resulting abilities to see at a glance what is relevant, to discern complex patterns and finer details, and to do so with minimal cognitive load, are hallmarks of expertise in all domains where humans attain remarkable levels of performance.

It is likely that this type of learning comprises a much bigger part of the learning task in many domains than has been understood in theoretical discussions of learning or implemented in methods of instruction. What is being discovered about PL has implications for learning and instruction that parallel what researchers in artificial intelligence have discovered, “that, contrary to traditional assumptions, high-level reasoning requires very little computation, but low-level sensorimotor skills require enormous computational resources” (http://en.wikipedia.org/wiki/Moravec's_paradox). An artificial intelligence researcher, Hans Moravec, elaborated this idea in what has come to be known as “Moravec’s Paradox” (Moravec, 1988):

Encoded in the large, highly evolved sensory and motor portions of the human brain is a billion years of experience about the nature of the world and how to survive in it. The deliberate process we call reasoning is, I believe, the thinnest veneer of human thought, effective only because it is supported by this much older and much more powerful, though usually unconscious, sensorimotor knowledge. We are all prodigious Olympians in perceptual and motor areas, so good that we make the difficult look easy. Abstract thought, though, is a new trick, perhaps less than 100 thousand years old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it.

In what follows, we will add elaborations of two kinds to Moravec’s Paradox. First, our Olympian perceptual abilities are astounding because they give us access to a great many of the abstract relations that underlie thought and action. “Sensorimotor knowledge” does not convey the scope and power of what perceptual mechanisms deliver. Not only is explicit abstract thinking possibly a newer evolutionary acquisition, but the work of abstraction is not exclusively the province of thinking processes alone. Much of thinking turns out to be seeing, if seeing is properly understood.

The second elaboration is that the evolutionary heritage that makes us perceptual Olympians involves not only fixed routines, but perceptual systems that change—that attune, adapt, and discover to optimize learning, problem solving, and complex task performance. These changes comprise a much larger component of learning and expertise than is usually understood in learning research. Such an understanding of PL has been even more conspicuously missing from the efforts to improve school learning and other formal instructional efforts.

In this chapter, we describe recent work on PL, with a particular focus on its relation to complex cognitive tasks. One important goal is to describe how PL relates to perception, cognition, and learning. Some of the domains in which we apply PL, such as mathematics, will seem distant from perception to many readers. Thus, the theoretical underpinnings of the effort deserve to be spelled out, and doing so may facilitate the understanding of current efforts and continuing progress in these areas. Making the basic connections here is important because the emerging understanding of PL has broad implications throughout the cognitive and neural sciences. Both understanding PL, and using it to improve learning, depend on coherent accounts of the relation between perception, cognition, and learning. A second aim of this chapter, building on the first, is to describe an emerging technology of PL that has many applications and offers the potential to address missing dimensions of learning and accelerate the growth of expertise in many domains. A large and growing research literature suggests that PL effects are pervasive in perception and learning, and that they profoundly affect tasks from the pickup of minute sensory detail to the extraction of complex and abstract relations in complex cognitive tasks. PL thus furnishes a crucial basis of human expertise, from accomplishments as commonplace as skilled reading to those as rarified as expert medical diagnosis, mathematical expertise, grandmaster chess, and creative scientific insight.

The article is organized as follows: In the next section, we consider the information-processing changes that are produced by PL. These have most often been examined in tasks that involve either low-level sensory discriminations or real-world tasks that obviously depend on perceptual discrimination (e.g. detecting pathology in radiologic images). Using the example of PL in mathematics learning, Section 3 extends PL to higher level symbolic cognitive tasks, in which PL has seldom been considered. Understanding the role of PL in such tasks requires a revised account of the relations of perception, cognition, and learning. In Section 4, we argue that the common conceptions of these processes and their relations do not provide a

satisfactory foundation for understanding high-level PL effects, primarily because they are based on outdated ideas about perception. Drawing on more recent views, we describe a framework for understanding PL components of high-level cognitive tasks that is rooted in the amodal and abstract character of perception itself. With this framework in hand, we consider more fully in Section V the applications of PL to instruction.



2. PERCEPTUAL LEARNING EFFECTS

Perceptual learning refers to experience-induced improvements in the pickup of information (E. Gibson, 1969). The fundamental observation is that perceptual pickup is not a static process. After an intensive period of research in the 1960s and a somewhat dormant period for two decades afterward, PL has become an area of concentrated focus in the cognitive and neural sciences. The relative neglect and occasional focus on PL in the history of learning research and its recent emergence have been described elsewhere, as have issues of modeling PL and understanding its neural bases (for a review, see Kellman & Garrigan, 2009). Another important question has been the relation between simple laboratory tasks involving PL and more complex, real-world tasks typically involving the extraction of invariance amidst variation; recent work suggests that all of these tasks partake of a unified learning process in which the discovery of relevant information and its selective extraction are key notions (Ahissar, Laiwand, Kozminsky, & Hochstein, 1998; Garrigan & Kellman, 2008; Li, Levi, & Klein, 2004; Mollon & Danilova, 1996; Petrov, Doshier, & Lu, 2005; Zhang et al., 2010). In the present discussion, we build on this recent work but do not revisit it. Here, we focus on the range of effects produced by the PL, before turning to more general issues of how these relate to basic notions of perception, cognition, and learning.

A wealth of research now supports the notion that, with appropriate practice, the brain progressively configures information extraction in any domain to optimize task performance. What are the changes involved? The list involves a variety of distinguishable effects that serve to improve performance. Kellman (2002) argued that these effects fall into two categories: *discovery* and *fluency* effects. Discovery effects involve finding what information is relevant to a domain or classification. Fluency effects involve coming to extract information with greater ease, speed, or reduced cognitive load. Table 4.1 summarizes some of the changes between novices and experts that occur from PL. Discovery effects

Table 4.1 Some Characteristics of Expert and Novice Information Extraction.

Discovery effects involve learning and selectively extracting features or relations that are relevant to a task or classification. *Fluency* effects involve learning to extract relevant information faster and with lower attentional load (see text).

	Novice	Expert
<i>Discovery Effects</i>		
Selectivity	Attention to irrelevant and relevant information	Selective pickup of relevant information Filtering/inhibition of irrelevant information
Units	Simple features	Larger chunks Higher-order relations
<i>Fluency Effects</i>		
Search type:	Serial processing	Increased parallel processing
Attentional load:	High	Low
Speed:	Slow	Fast

include the fundamental idea of *selection* (Gibson, 1969; Petrov et al., 2005): We discover and pick up the information relevant to a task or classification, ignoring, or perhaps inhibiting (Kim, Imai, Sasaki, & Watanabe, 2012; Wang, Cavanagh & Green, 1994) available information that is irrelevant. We come to extract information in larger chunks, forming and processing higher-level units (Chase & Simon, 1973; Goldstone, 2000). Most profoundly (and mysteriously), we come to discover new and often complex relationships in the available information to which we were initially insensitive (Chase & Simon, 1973; Kellman, 2002). These discovery processes are pervasive in early learning. When a child learns what a dog, toy, or truck is, this kind of learning is at work. From a number of instances, the child extracts relevant features and relations. These allow later recognition of previously seen instances, but more important, even a very young child quickly becomes able to categorize *new* instances. Such success implies that the learner has discovered the relevant characteristics or relations that determine the classification. As each new instance will differ from previous ones, learning also includes the ignoring of irrelevant differences.

Fluency effects refer to changes in the efficiency of information extraction. Practice in classifying leads to fluent and ultimately automatic processing (Schneider & Shiffrin, 1977), where automaticity in PL is defined as the

ability to pick up information with little or no sensitivity to attentional load. As a consequence, perceptual expertise may lead to more parallel processing and faster pickup of information.

The distinction between discovery and fluency effects is not always perfectly clear. For example, becoming selective in the use of information (a discovery effect) increases efficiency and improves speed (fluency effects). It does seem, however, that there are clear cases of each category. In one of the earliest relevant studies, [Bryan and Harter \(1899\)](#) reported that telegraph operators learning to receive Morse code reached plateaus in performance, but that continuing practice while at a plateau appeared to pave the way for substantial new gains in performance. Their interpretation is that the eventual improvements in performance came from automaticity—operators coming to extract the same information with less cognitive load, ultimately enabling them to discover more complex relations in the input. This interpretation is consistent with a relatively pure fluency improvement, that is, with practice at a certain point not changing the information being extracted, but allowing its extraction with reduced attentional load ([Schneider & Shiffrin, 1977](#)). The continuing cycle of discovery and fluency described by Bryan and Harter—discovery leading to improved performance, followed by improved fluency, leading in turn to higher level discovery—may be the driver of remarkable attainments of human expertise in many complex tasks.



3. PERCEPTUAL LEARNING IN MATHEMATICS: AN EXAMPLE

There is a common view about the relation of perception and cognition. In a hierarchy of cognitive processes, perception is typically considered “low-level,” where “higher” cognitive processes encompass categorization, thinking, reasoning, etc. Eleanor Gibson, who pioneered the field of PL, thought of it as a pervasive contributor to expertise, giving examples as varied as chick sexing, wine tasting, map reading, X-ray interpretation, sonar interpretation, and landing an aircraft. Even these examples, however, are mostly confined to tasks where the major task component is classifying perceptual inputs based on subtle kinds of information. For most of these examples, one might still maintain a notion of perception as handing off results of basic feature detection, which then become the raw material for conceptual analysis, cognitive inferences, and high-level thinking.

Recent work, however, indicates that PL is strongly involved even in very high-level cognitive domains, such as the learning and understanding of mathematics (Kellman, Massey, & Son, 2009; Landy & Goldstone, 2007). Learning in these domains involves a variety of cognitive processes, but attaining expertise depends substantially on pattern recognition and fluent processing of structure, as well as mapping across transformations (e.g. in algebra) and across multiple representations (e.g. graphs and equations). In fact, given conventional instruction, the PL components of expertise may be disproportionately responsible for students' difficulties in learning (Kellman et al., 2009). Although this research area is relatively new, findings indicate that even short PL interventions can accelerate the fluent use of structure, in contexts such as the mapping between graphs and equations (Kellman et al., 2008; Silva & Kellman, 1999), apprehending molecular structure in chemistry (Wise, Kubose, Chang, Russell, & Kellman, 2000), processing algebraic transformations, and understanding fractions and proportional reasoning (Kellman et al., 2009).

The structures and relations that are relevant to PL in these domains are more abstract and complex than what we normally think of as being processed perceptually. As an example, Kellman et al. (2009) studied algebra learning using a perceptual learning module (PLM) designed to address the seeing of structure in algebra. Participants were 8th and 9th graders at midyear in Algebra I courses. Students at this point in their learning show a characteristic pattern. Given simple equations to solve, such as $x + 4 = 12$, accuracy is high, with an average across participants of around 80% correct solutions. Remarkably, however, students at this stage take an average of about 28 s per problem! This pattern suggests that conventional instruction does a good job of addressing the declarative and procedural aspects of solving algebraic equations. Students know they should “get x alone on one side,” and “do the same operation to both sides of the equation,” and they were able to accomplish these goals with high accuracy. Their response times, however, suggest that we may underestimate the seeing problem in algebra learning. Someone with much greater experience looks at $x + 4 = 12$ and sees the answer at a glance. This kind of ability can reach higher and higher levels, supporting greater expertise, as illustrated in this example:

$$\mu = \frac{(4\phi - 2\phi\psi)}{(2 - \psi)(2\phi)}.$$

Given that this is a single equation with two unknowns, one might think at first glance that the problem does not permit a numerical solution for μ , but a more practiced observer may easily see that the equation permits easy simplification, and $\mu = 1$. In this case, even the relative unfamiliarity of the symbols used may make the seeing problem harder. Without changing anything mathematical, compare

$$m = \frac{(4x - 2xy)}{(2 - \gamma)(2x)}.$$

If this equation still has you reaching for pen and paper, seeing the structure may be better illustrated in this simpler version:

$$m = \frac{(x - xy)}{(x)(1 - \gamma)}.$$

These examples all involve the distributive property of multiplication over addition. However, being able to enunciate this property would not produce fluent recognition of the distributive structure. Conceptually, and even computationally, these examples are all very similar, but you may have noticed the relevant structure more easily in one case than another. Improved encoding of relevant structure and potential transformations in equations is a likely result of PL, one that is difficult to address in conventional instruction.

Following this kind of intuition, we developed our *Algebraic Transformations PLM* in order to apply PL methods to improve students' pattern processing and fluency in algebra. We developed a classification task in which participants viewed a target equation or expression and made speeded judgments about which one of a set of possible choices represented an equivalent equation or expression, produced by a valid algebraic transformation. A key goal of this PLM was to contrast the declarative knowledge components (facts and concepts that can be verbalized) with the idea of "seeing" in algebra. The goal was to get students to see the structure of expressions and equations, and relations among them, in order to use transformations fluently.

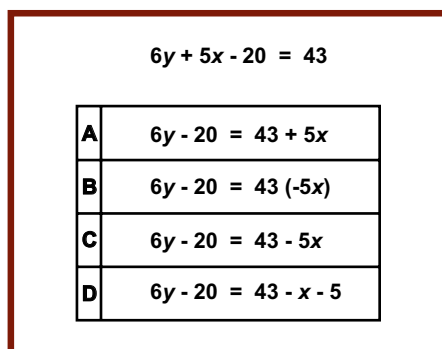
In the *Algebraic Transformations PLM*, we did not ask students to solve problems. Instead, we devised a classification task that exercised the extraction of structure and the seeing of transformations. On each trial, an equation appeared, and the student had to choose which one of several options below was a legal transformation. An example is shown in [Figure 4.1](#). In addition to testing whether practice in the PLM improved accuracy and

fluency in recognizing transformations, we also examined whether students would be able to transfer learning to solving algebraic equations.

This study was carried out with forty-two 8th and 9th grade students at midyear in an Algebra 1 course. Students participated in two 40-min learning sessions using the *Algebraic Transformations PLM*. On each trial, they were shown a target equation and were asked to select which of four choices could be correctly derived by performing a legal algebraic transformation on the target. Students were given feedback after each trial indicating whether or not they had chosen the correct answer. Incorrect answers were followed by an interactive feedback screen in which students' attention was focused on the relevant transformation.

The task that formed the core of the PLM—matching an equation to a valid transformation—is directly useful to development of pattern recognition and skill in algebra. The PLM produced dramatic gains for virtually all students on this task, with accuracy changing from about 57% on initial learning trials to about 85% at the end of PLM usage, and response times per problem reduced by about 55%, from nearly 12 s per problem to about 7 s, suggesting the development of fluency in processing symbolic structure of equations.

Perhaps more remarkable was the transfer to actual algebra problem solving. Although students did not receive any practice in solving equations during the learning phase, the relatively brief intervention aimed at seeing transformations produced a dramatic reduction in the post-test equation solving time—from about 28 s per problem to about 12.5 s per problem (Figure 4.2, right panel). A delayed post-test showed that these gains were lasting: The average solving time was actually slightly faster than in the immediate post-test when tested after a 2-week interval. There was



$6y + 5x - 20 = 43$	
A	$6y - 20 = 43 + 5x$
B	$6y - 20 = 43 (-5x)$
C	$6y - 20 = 43 - 5x$
D	$6y - 20 = 43 - x - 5$

Figure 4.1 *Example of a Problem Display in the Algebraic Transformations PLM* (see text).

also some indication that accuracy in equation solving, already high at pretest, received some benefit in the delayed post-test (Figure 4.2, left panel).

The idea that mathematical understanding has an important PL component may seem counterintuitive, for many reasons. If perception is about properties such as brightness, color, the orientation of edges, or even the locations of objects and surfaces, how is this relevant to a mathematics class? These perceptual contents might at best serve up the occasional concrete example, but they hardly encompass mathematical ideas. On traditional views, most of mathematical thinking, and the instructional methods used to teach it, involve declarative knowledge and procedures. Perception may serve the banal role of allowing the student to see the markings on the chalkboard, but the processing of mathematical ideas must surely be farther up in the cognitive hierarchy! There would seem to be a gap between the basic and concrete information furnished by the senses and the abstract conceptual content of mathematics. The simple difference between the level or types of information that perception is presumed to furnish and what is required for abstract thinking seems a formidable obstacle to the kind of connection we are making here.

But it is not the only obstacle. Mathematics has inherently symbolic aspects. The symbols in an equation have an *arbitrary* relation to the ideas they represent. Unlike the functional properties of objects and events in the world, the meanings of mathematical ideas would seem remote from

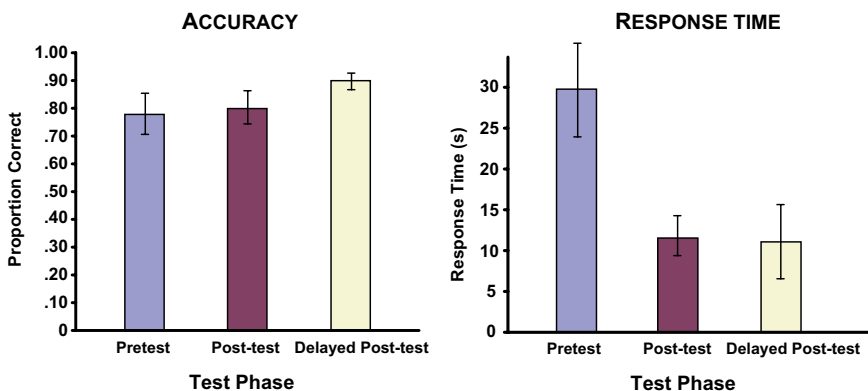


Figure 4.2 Results of Algebraic Transformations PLM Study for the Transfer Task of Solving Algebraic Equations. Data for pretest, post-test, and delayed post-test are shown for accuracy (left panel) and response time (right panel). Error bars indicate ± 1 standard error of the mean (Adapted from Kellman, Massey & Son, *TopiCS in Cognitive Science*, 2009; *Cognitive Science Society, Inc.*, p. 14). For a color version of this figure, the reader is referred to the online version of this book.

stimulus information reaching perceptual systems. Moreover, much of the expertise conferred by PL may be implicit (e.g. try describing to a stranger how you recognize your sister's voice on the telephone), whereas mathematics is in many respects an extremely explicit discipline. Steps must be justified and proofs must be offered. Even assuming the relevance of PL to complex tasks, one might still wonder about the application to symbolic, explicit domains such as mathematics.

Many of these objections have straightforward answers. Even if they involve symbolic content, mathematical representations pose important information extraction requirements and challenges. Characteristic difficulties in mathematics learning may directly involve issues of discovery and fluency aspects of PL. A number of studies indicate the role of PL in complex cognitive domains, such as mathematics (Kellman et al., 2009; Landy & Goldstone, 2007; Silva & Kellman, 1999), language or language-like domains (Gomez & Gerken, 1999; Reber, 1993; Reber & Allen, 1978; Saffran, Aslin, & Newport, 1996), chess (Chase & Simon, 1973), and reading (Baron, 1978; Reicher, 1969; Wheeler, 1970). Some have asserted that in general, abstract concepts have crucial perceptual foundations (Barsalou, 1999; Goldstone, Landy, & Son, 2008; Prinz, 2004).

The extensive use of tangible representations in mathematics, science, and other abstract conceptual domains is also a bit of a giveaway. Hardly two steps into considering a complicated problem in mathematics, science, economics, or other quantitative disciplines we construct a graph or a diagram, if not several. One's facility in dealing with these representations obviously changes with experience, in obscure ways that go beyond being able to explain the basics of how the diagram represents information. We seem to grapple with complex ideas in mathematics and science by using spatial, configural, and sometimes temporal structures (i.e. simulations) that draw on representational capacities rooted in our perceptions of spatial and temporal structure in the world. A graph of the change of world temperature over time is a spatial object, and the patterns therein are comprehended by grasping spatial relations, although neither temperature nor time is a spatial notion. Reliably accompanying the use of these structures and representations are powerful, general capacities to learn to detect relations and become able to fluently select information that is important within a domain: Perceptual learning.

Still, we are stuck with the first objection. Perception, as commonly understood, just seems to be at the wrong level for explaining comprehension in mathematics. Maybe the connection is intended as some of kind of metaphor. If one conceives of perception as consisting of separate sense

modalities, then what we obtain through vision must somehow be built from sensory experiences of brightness and color. In addition, we are presumably extracting sequences and combinations of loudness and pitch. In algebra class, one should listen to the teacher's voice and look at the blackboard, but surely algebra is not about arrays of color, brightness, loudness, or pitch.

Later in this chapter, we will have more to say about PL technology and the potential for radically improving learning by integrating methods that accelerate PL with conventional instruction. For now, however, we focus on what appears to be most perplexing in our example of PL in complex cognitive domains. If it is surprising that changing the perceiver can be the key to advancement in domains such as mathematics, it is because there is work to do in clarifying the relation of perception to learning and cognition. This is the focus of the next section.



4. PERCEPTION, COGNITION, AND LEARNING

Continuing scientific progress and practical applications of PL will be facilitated by a better understanding of the relations between perception, cognition, and learning. One might assume that these relations are well understood, but in fact they are not. A primary reason is that progress in understanding perception in the past several decades necessitates a rethinking of some of these relations, invalidating some ways of thinking and paving the way for new insights.

As we mentioned above, commonly held views of perception would suggest that the products of perceiving are too low level to have consequences for abstract thinking and learning. Thus, before the last few years, if someone suggested a role for perception in learning mathematics, it would involve using shaded diagrams to illustrate fraction concepts or manipulatives that might allow learners to have some concrete realization of adding numbers. These applications are quite different from the idea of a general learning mechanism by which learners progressively change the way they extract structure and relations from symbolic equations, or gain competency in mapping structure across differing mathematical representations, or come to selectively attend to important relations, rather than irrelevancies, in a measurement problem.

In recent years, there have been trends in cognitive science arguing for a close relation between perception and cognition. This work includes empirical findings that implicate perceptual structure as being involved in processing abstract ideas (Landy & Goldstone, 2007) and other research indicating modal sensory activations accompanying cognitive tasks such as

sentence verification (van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008). There have been accompanying theoretical proposals that suggest that high-level cognition depends fundamentally on perception, including the ideas of perceptual symbol systems (PSS; Barsalou, 1999) and the notion of embodied cognition. We believe that these accounts share important elements, and all are an improvement on an earlier, implicit general view of perception being detached from thinking.

In our view, however, none of these efforts provides a suitable basis for understanding the relation of perception and PL to the rest of cognition and to complex learning domains. As a result, the situation is confusing. We have found this to be especially troubling in terms of connecting emerging findings in PL and PL technology in instruction with conventional ideas of cognition, teaching, and learning. The reason is that neither the older assumptions about how these relate nor most recent proposals in cognitive psychology provide a coherent basis for understanding the relation of PL to cognition in general. We briefly discuss some of these views and their problems before describing a more coherent, as well as simpler, account, one grounded in a contemporary understanding of perception.

4.1. The Classical View of Perception

In classical empiricist theories of perception and perceptual development, widely shared for several centuries by many philosophers and psychologists, all meaningful perception (e.g. perception of objects, motion, and spatial arrangement) was held to arise from initially meaningless sensations. Meaningful perception was thought to derive from associations among sensations (e.g. Berkeley, 1709/1910; Locke, 1690/1971; Titchener, 1902) and with action (Piaget, 1952). In this view, all of perception is essentially a cognitive act, constructing meaning by associating sensations and connecting them with previously remembered sensations. A modern version of this view, widely shared in cognitive psychology, is satirized in a famous information-processing diagram in Ulric Neisser's book *Cognition and Reality* (Neisser, 1976), in which an input labeled "retinal image" is connected by arrows to successive boxes labeled "processing," "more processing," and "still more processing."

This view of perception came with its own view of PL. Essentially, on this view, *all* meaningful perception is a product of learning. Inferring the motion of an object from sensations encoded at different positions and times, or understanding the three-dimensional shape of an object by retrieving previously stored images gotten from different vantage points involve meaningless sensations combined with associative learning processes (e.g.

Locke, 1690/1971), or unconscious inference processes working on current and previously stored sensations (Helmholtz, 1864/1972).

4.2. Perceptual Symbol Systems

There have been clear trends among cognitive researchers to connect perception more closely to other cognitive processes or to uncover perceptual influences in cognitive tasks. Particularly influential has been the work of Barsalou on “perceptual symbol systems” (PSS). PSS comprise proposals to account for a number of important phenomena, including well-known difficulties of specifying formal, context-free criteria of inclusion in conceptual categories (e.g. what makes something a cat); the apparently dynamic, variable aspects of representations; and the engagement of cortical areas involved with perception during cognitive tasks.

According to PSS, the idea of nonperceptual, abstract thought does not really exist. Even our most abstract ideas are attained by reference to stored perceptual encodings. As Barsalou (1999) explains,

...abstract concepts are perceptual, being grounded in temporally extended simulations of external and internal events. (Barsalou, 1999, p. 603)

Specifically, when we think of an apparently abstract idea, that processing consists of running a “simulation,” which consists of “re-enactment in modality-specific systems”:

The basic idea behind this mechanism is that association areas in the brain capture modality-specific states during perception and action, and then reinstate them later to represent knowledge. When a physical entity or event is perceived, it activates feature detectors in the relevant modality-specific areas. During visual processing of a car, for example, populations of neurons fire for edges, vertices and planar surfaces, whereas others fire for orientation, colour and movement. The total pattern of activation over this hierarchically organized distributed system represents the entity in vision (e.g. Zeki 1993; Palmer, 1999). Similar distributions of activation on other modalities represent how the entity feels and sounds, and the actions performed on it. (Barsalou, 2003, p. 1179)

Barsalou contrasts this view with what he sees as the more typical view in cognitive science that information gotten through perception is “transduced” into amodal representations, where

... an amodal symbol system transduces a subset of a perceptual state into a completely new representation language that is inherently nonperceptual. (Barsalou, 1999, p. 577)

We believe that Barsalou and others have identified a key problem—a perceived disconnect between information processing involving most

cognitive processes and perception. The general idea that these are more closely coupled than often believed is consistent with considerable evidence and has opened up some important issues in these fields.

4.3. Problems for Understanding Perceptual Learning

Regrettably, both the classical view and more recent proposals about the relation between perception and cognition make poor foundations for understanding current approaches to and results of PL. Mathematics seems much more abstract than perception. Consider the applications of PL to mathematics that we described above. On the classical view, it is hard to relate the abstract structures in mathematics to the aggregates of sensations that are the harvest from perceiving. Mathematics seems to be the province of higher-level reasoning, not perception.

The situation is somewhat reversed from Barsalou's PSS view. Here, it is claimed that abstract ideas do not really exist off by themselves; what we think of as abstract thought really consists of activations of modality-specific features. On this account, all mathematics would be inherently perceptual. It is hard to see how it would be abstract, however. If the input contents are all modality specific, what is mathematics? Is mathematics visual? Is it auditory? Tactile? Mathematics does not really seem to be any of those things. From the PSS account, it could be argued that thinking about a mathematical idea involves running certain re-enactments of particular perceptual experiences. These are likely multimodal; they could have inputs from different modalities such as the sound of your teacher's voice in algebra class or the chalkmarks on the blackboard, or the feel of your pencil in your hand. Thinking about particular ideas in particular contexts would involve re-enacting (simulating) different subsets of stored perceptual records.

Both this approach and the classical approach have massive problems with *abstraction* and *selection*. Consider a student who is mastering the concept of slope in a PLM involving graphs, equations, and word problems. The student's task is to map a problem represented in one format, such as a graph, to the same mathematical structure as it is expressed in either an equation or a word problem. The student learns to extract a general idea that applies to new contexts, as well as structural invariants specific to representational types (Kellman, Massey & Son, 2009). In a graphic representation, the understanding of slope emerges as involving spatial directions: A positively sloped function increases in height from the left to the right; steeper increases show larger slopes, and so on. From mapping word problems onto graphs, the deeper understanding emerges that a positive slope

involves increases in one quantity as another quantity increases. As water is heated, a rising temperature over time implies a positive slope. One could well recall an experience of boiling water when one thinks about slope, but that would not help without some mechanism of specifying which parts of that experience constitute slope. The slope concept can apply to boiling water but is not about boiling water. It has been argued that the PSS framework involves insurmountable problems in that rerunning various perceptual records provides no mechanism for selecting a particular idea (Landau, 1999; Ohlsson, 1999). The problem is especially severe when the idea is an abstract one, such as slope. Meaningful learning here would involve a student being able to apply slope to novel situations (e.g. knowing what it would mean if there were a negative slope relating number of business startups to interest rates). It is hard to fathom how this understanding of a novel case could come from rerunning subsets of prior modality-specific activations. As Ohlsson (1999) puts it,

A closely related difficulty for Barsalou's theory is that the instances of some concepts do not share any perceptible features. Consider furniture, tools, and energy sources. No perceptible feature recurs across all instances of either of these categories. Hence, those concepts cannot be represented by combining parts of past percepts. (Ohlsson, 1999, p. 630)

PL in complex cognitive domains leads to selective extraction and fluent processing of abstract relations, such as slope. From transactions with individual cases, learners come to zero in on the properties, including abstract relations, that underlie important classifications. The process is PL, as it changes the way information is extracted. The learning is highly selective; selection is so fundamental to PL that Gibson (1969) used “differentiation learning” as a synonym for PL. Finally, the properties learned are abstract. Whether in chess, speech recognition, chemistry, or mathematics, PL often leads to selective, fluent extraction of relational and abstract information (Kellman & Garrigan, 2009).

Traditional views of perception and recent proposals regarding perception and cognition, such as PSS, do not appear to offer reasonable ways of understanding these aspects of PL. How can we understand them? To begin with, the answer can be found in a better understanding of perception itself.

4.4. The Amodal, Abstract Character of Perception

Both the classical view of perception and recent attempts to connect perception and cognition are hampered by a failure to understand the amodal, abstract character of perception.

Research and theory in perception over the past several decades have made it clear that perceptual systems are sensitive to complex relations in stimulation as their inputs, and they produce meaningful descriptions of objects, spatial layouts, and events occurring in the world (J. Gibson, 1966, 1979; Johansson, 1970; Marr, 1982; Scholl & Tremoulet, 2000). Most perceptual mechanisms develop from innate foundations or maturational programs and do not rely on associative learning to provide meaningful perception of structure and events (for a review, see Kellman & Arterberry, 1998). Many structural concepts that might earlier have been considered exclusively cognitive constructs have been shown to be rooted in perceptual mechanisms. Some of these include causality (Michotte, 1963), animacy (Johansson, 1973), and social intention (Heider & Simmel, 1944; Runeson & Frykholm, 1981; Scholl & Tremoulet, 2000).

These features of perception are difficult to reconcile with a shared assumption of classical views, PSS, and some other approaches that the products of perceiving are sets of sensory activations that are modality specific—that is, unique to particular senses. In PSS, for example, the definition of perceptual symbols requires that they be modality specific, consisting of records of “feature activations” (Barsalou, 1999, 2003). This approach to representation, according to Barsalou, replaces the amodal symbols common in other cognitive modeling, resulting in the view that there may be no truly abstract, amodal symbols at all.

Any approach of this sort is difficult to reconcile with the fact that most of the perceptual representations that are central to our thought and action have a distinctly nonsensory character. For example, as the Gestalt psychologists pointed out almost 100 years ago, the perceived shape of an object is something quite different from the collection of sensory elements it activates (Koffka, 1935). The problems with obtaining the products of perception from aggregates of sensory activations are well known (J. Gibson, 1979; Koffka, 1935; Kellman & Arterberry, 1998; Landau, 1999; Marr, 1982; Nanay, 2010; Ohlsson, 1999).

The solution of how to connect perception with abstract thought is not that abstract thought consists of simulations of sensory feature activations but that perception *itself* is amodal and abstract. The terms “modal” and “amodal” were in fact introduced in perception research by Michotte, Thines, and Crabbe (1964) with regard to these issues. Michotte et al. use *both modal* and *amodal* to refer to perceptual phenomena. In his classic work on visual completion, modal completion refers to cases in which the visual system fills in information that includes sensory properties, such

as brightness and color, whereas amodal completion refers to filling-in in which the object structure is represented perceptually but there is an absence of sensory properties. (The latter happens, for example, when an object is seen as continuing behind a nearer occluding object.) Michotte's view, supported by extensive research, is that both kinds of filling-in are accomplished by perceptual mechanisms, not processes of reasoning or cognition (Kanizsa, 1979; Keane, Lu, Papatomas, Silverstein, & Kellman, 2012; Michotte et al., 1964). In fact, both kinds of filling-in appear to depend on the same perceptual mechanisms (Kellman & Shipley, 1991; Kellman, Garrigan, & Shipley, 2005; Murray, Foxe, Javitt, & Foxe, 2004). In general, visual perception of ordinary surfaces and objects results in representations of complete objects and continuous surfaces, even when many parts of these are not represented in local sensory input due to occlusion or camouflage (Kanizsa, 1979; Kellman & Shipley, 1991; Michotte et al., 1964; Palmer, Kellman, & Shipley, 2006).

The issue here may be in part terminological. Barsalou (1999, 2003) defines perceptual symbols in general as necessarily "modal," and contrasts these with the nonperceptual or "amodal" symbols. His explication of modal perceptual symbols includes being the property of a single sense and being "analogical," in that such symbols are "represented in the same systems as the perceptual states that produced them. The structure of a perceptual symbol corresponds, at least somewhat, to the perceptual state that produced it" (Barsalou, 1999). One could explore the idea that Barsalou may be giving the terms "modal" and "amodal" new meanings and therefore there is no conflict with Michotte's ideas. On this view, anything vision does is "modal" because vision is one sense, as distinguished, for example, from audition. The nonsensory phenomena of visual object and surface perception, and so on, would simply be modal under these new definitions.

The different use of terms is accompanied by a difference in concept, however. The problem is clear in the proposals that perceiving an object consists of feature activations, such as neurons for edges, vertices, orientation, color, etc., and that "The total pattern of activation over this hierarchically organized distributed system represents the entity in vision." Barsalou's view is in many ways remarkably close to classical views of sensation and perception, as he notes (Barsalou, 1999, p. 578).

In the field of perception, Michotte's ideas were incorporated into the more comprehensive ecological, information-based theories of J. J. Gibson (1966, 1979). Gibson made the case that perceptual mechanisms have evolved to be sensitive, not to simple, local stimuli, but to higher order relations

(invariants) in stimulation that correspond to important environmental properties or important events involving the perceiver and the environment. Much of the important information is not even present in a particular, momentary sensory array (image). For example, variables in optic flow—the continuously transforming projection of the environment onto the eyes—specify the direction of travel of a moving observer, as well as the layout of surfaces ahead (Gibson, 1979; Warren & Hannon, 1988). In general, Gibson embraced the idea of perception, at least its most functionally important aspects, as “sensationless.”

An example of the extraction of complex relations by perceptual mechanisms to produce descriptions of high-level, abstract properties may help to make this idea intuitive. Johansson (1973) placed small lights (“point lights”) on the joints of a person, and filmed the person walking in the dark. When viewed by a human observer, there is a compelling and automatic percept of a person walking. Such displays may also convey information about gender or specific individuals. Many more complex events involving so-called *biological motion* have been shown to be quickly and effortlessly perceived, including dancing and jumping.

Any static view of the dots used in these displays conveys only a meaningless jumble. Moreover, dot displays, in momentary images or in motion, do not at all resemble any stored images (or sets of feature activations) we may have of actual walking (or dancing) persons. All the basic sensory features in these displays are, upon first presentation, brand new. Moreover, the observer represents perceptually a walking person and encodes in a durable fashion almost nothing about positions of particular dots in momentary images, or dot trajectories, that comprised the animation sequence. The fact that observers uniformly and automatically perceive meaningful persons and events in these displays indicates that our normal encoding of persons and events in the environment includes abstract relations of high complexity.¹ All these observations illustrate crucial and general aspects of perception: We do register sensory elements (and feature activations), but we do so as part of processes that extract complex and abstract relations relevant to detecting ecologically important properties of objects and events. It is these properties that are encoded; the basic sensory features are transient, quickly discarded, and, apart from the relations in which they participate, quite irrelevant to perception. These ideas that perceptual systems utilize complex relational

¹ They are complex enough that scientists who study computational vision have not yet been able to produce algorithms to approximate human performance in perceiving structure from point-light displays.

information as inputs and produce abstract, amodal representations as outputs are shared by virtually all contemporary ecological and computational work in perception (Hochberg, 1968; Kellman & Arterberry, 1998; Marr, 1982; Shepard, 1984; Pizlo, 2010) and are not subjects of serious dispute.

We should note specifically that this view of the outputs of perception as amodal, meaningful abstractions applies even to seemingly simple cases of perception. The idea that we could represent some object in the world, say, a car, in terms of sets of feature detectors activated in sensory areas, constitutes a vast and misleading simplification. It is true that early cortical areas in the visual system contain orientation-sensitive units that respond to retinally local areas of oriented contrast. So it may seem straightforward to assume that activations of such cells could represent the oriented edges of a car that we see. But this is a misunderstanding. The perceived orientation of an edge of a car in the world is actually the result of complex computations accomplished by perceptual mechanisms; it is not a readout of the outputs of early orientation-sensitive units. One reason is that capturing information about an edge in the world requires utilizing relations among many different orientation-sensitive units of different local orientations and scales (e.g. Lamme & Roelfsema, 2000; Sanada & Ohzawa, 2006; Wurtz & Lourens, 2000). Another problem is that the early neural units in vision encode two-dimensional orientations on the retina, not the three-dimensional orientations in space needed in our perceptual representations (for discussion, see Kellman et al., 2005). The most general version of the problem here, however, is that the word “orientation” means different things for the “feature detectors” of the basic vision scientist and the object “features” needed in cognitive models. The former are invariably *retinal*, meaning that the orientation-sensitive units in V1 that get activated depend on the orientation and position of contrast on the retina of the eye. This position and orientation information typically changes several times a second,² as it depends crucially on the position of the eyes in the head, the head on the body, and the body in the world. Thus, the correspondence between the orientation of an edge in the world and which orientation-sensitive units are firing in the brain is haphazard. Complex relations in the activities of orientation-sensitive units allow us to encode properties of objects in a world-centered coordinate system, but there is no reason to believe that we encode into any

² Even identifying an early cortical unit with a single retinal orientation is an oversimplification. In fact, early cortical units in vision have complex response profiles that include changes in their orientation sensitivity over periods <100ms, and they are sensitive to many other influences of context (Lamme & Roelfsema, 2000; Ringach, Hawken, & Shapley, 2003).

lasting form of storage any sensory records of the momentary activations of neurons in early cortical areas. In fact, there are reasons to believe that early activations of “feature detectors” in early visual processing are not accessible to even momentary conscious awareness (e.g. Crick & Koch, 1995; He & MacLeod, 2001) and cannot be accessed by learning mechanisms (Garrigan & Kellman, 2008). In short, the perception of a simple environmental property, such as the edge of a car, is a complex abstraction, based on relational information; the relations of this abstraction to the outputs of populations of detectors, such as the orientations signaled by neuron in early visual areas, are highly variable; and the latter are not preserved in any accessible outputs of the process. Elementary activations in sensory areas are not the elements of perceptual representations—not even the seemingly simple ones, such as orientation or color (see Zeki, Aglioti, McKeefry, and Berlucchi (1999) for a parallel argument regarding color). We would go so far as to say that the term “feature detector” has proven to be an unfortunate choice in sensory neuroscience. When construed to mean that early neural units signal the features of objects, surfaces or events in the world, it is a misunderstanding.

The transition from a view of perceptual representations as some kind of energy imprint on the sensory surfaces to a view of these representations as amodal and abstract parallels the developments in other sciences. Commenting on the ways in which quantum mechanics had changed conceptions of matter from continuous and concrete to something much more abstract, the philosopher Bertrand Russell put it: “It has begun to seem that matter, like the Cheshire Cat, is becoming gradually diaphanous and nothing is left but the grin, caused, presumably, by amusement at those who still think it is there.” The lingering view of perception, and the representations gotten from perception, as being fundamentally about local sensory activations is just like this. In contemporary views, sensory activations provide a medium from which perceptual mechanisms extract informative relations in order to represent in abstract fashion ecologically important structures of objects and events. The sensory Cheshire cat has proved similarly diaphanous, leaving nothing but the grin. (Even the grin, if we recall it from having seen a picture earlier, is an abstract structure, not an array of sensations).

4.5. The Selective Character of Perception

Earlier, we noted that selection is a key principle in PL. It is a crucial characteristic, because organisms are surrounded at any moment by a wealth of stimulation. The tasks they need to perform require highly selected subsets of this information, and sometimes require discovery of complex, subtle,

and abstract properties and relations. Moreover, we have limited immediate processing capacities, such that cognitive load is a major constraint on performance in most tasks, and conspicuously so in learning (Paas & van Merriënboer, 1994; Schneider & Shiffrin, 1977; Sweller & Chandler, 1991). Selective apprehension of information and improvements in fluency (speed, chunking, and automaticity or reduced load) with practice are both primary mechanisms by which humans cope with these limitations and major determinants of expertise in most domains.

As we noted above, information selection in classical views would be hard to accomplish because pickup was based on sensations, not information. Both fashioning abstract ideas out of associations of sensations and altering the information extraction process with experience are hard to fathom from this starting point. The situation is different but equally problematic from the PSS perspective. Again, if records of feature activations for whole episodes are what is picked up and what is stored, selection and isolation of invariants or distinguishing features pose an unsolved problem.

Fortunately, the problem is much more easily handled in contemporary views of perception, in which selection plays an important and intrinsic role. As we have seen, selective computation of abstract properties, from simple ones such as edge orientation in space, to more complex ones such as shape or sets of motion relationships that specify objects, surfaces or events, is a fundamental characteristic of perception (J. Gibson, 1979) and appears to be presumed by learning mechanisms as well (Garrigan & Kellman, 2008).

4.6. Common Amodal Perceptual Representations for Thought, Action, and Learning

What is the format of perceptual representations? A holdover from traditional theories is that information that comes in through sight is encoded in a visual representation, information gotten through hearing is encoded in an auditory representation, and so on. Products of different senses, if stored in separate encodings, would have to be subject to endless associations and calibrations to achieve even the simplest results in representing the world. When you perceive a bird that both squawks and flaps its wings, your brain would require complicated transactions to relate the squawking in the auditory world to the flapping in the visual world.

The idea that perceptual information must be saddled with the fragmentation of a separate visual world, an auditory world, a tactile world, etc. came originally from the obvious fact that we use different sense organs to pick up information, the unique sensations that characterize each sense,

and from the assumption that the contents of perception were aggregates of these sensations. Such an account leads irrevocably to the idea of separate representations in the separate senses, along with the need for associative or inference processes that have to be used to connect them.

If, as is now recognized, separate sensory input channels furnish more abstract information about structure in the world, it would make sense that, at least to some degree, these outputs converge into a common representation. Unlike the ideas that perception is amodal and information based, this idea is not yet a consensus view; it is still common for researchers to discuss multisensory integration or amodal representations in distinct senses (Nanay, 2010; Pouget, Deneve, & Duhamel, 2002). Yet, along with the obvious functional utility of having perceptual descriptions in a common amodal representation, there is now considerable evidence for early and intrinsic connections across the senses (Falchier, Clavagnier, Barone, & Kennedy, 2002; Knudsen & Knudsen, 1985; Meltzoff & Moore, 1989; Spelke, 1987; Stein & Meredith, 1993; Wertheimer, 1961).

Some neurophysiological evidence directly implicates encoding of amodal properties, such as location, apart from particular sensory channels. Knudsen (1982) discovered cells in the optic tectum of barn owls that respond to locations in space, whether specified auditorily or visually. This is direct evidence for a system encoding information about space and time, into which sensory channels feed, rather than a set of separate sensory representations. Much recent work in a variety of mammalian species also suggests that the brain is wired to connect the sensory input channels much earlier than was previously understood. Even early cortical areas, such as V1 and A1, that have been considered exclusively involved with one sense, have been shown to have multisensory influences (Falchier et al., 2002; Ghazanfar & Schroeder, 2006; Stein & Meredith, 1993). Stein and Stanford (2008, p. 263), in reviewing an extensive neurophysiological literature, conclude that "...evidence of early multisensory convergence raises fundamental questions about the sensory-specific organization of the cortex" and "These observations question whether there are any exclusive, modality-specific cortical regions and, thus, whether it is worth retaining designations that imply such exclusivity."

A wide variety of evidence and argument supports the idea that to support learning, thought and action, perceptual descriptions must involve a common, amodal representation, rather than merely modality-specific records (Ernst & Banks, 2002; Klatzky, Wu, & Stetten, 2008; Lehar, 1999; Stoffregen & Bardy, 2001).

4.6.1. Embodied Cognition

Another important question is whether these representations must always be tied to *actions*. The idea of embodied cognition is a relatively recent set of ideas that suggests a close relationship between perception, cognition, and action. Like Barsalou's PSS approach, embodied cognition views tend to deny the idea of abstract cognitive representations separate from episodes of perceiving and acting. Thelen (2000) expresses the idea this way:

To say that cognition is embodied means that it arises from bodily interactions with the world and is continually meshed with them. From this point of view, therefore, cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capabilities ... (p. 5)

One issue in evaluating embodied cognition views is that there are a variety of them. Wilson (2002) has identified at least six possible basic claims of embodied cognition. They are:

1. Cognition is situated. Cognitive activity takes place in the context of a real-world environment, and it inherently involves perception and action.
2. Cognition is time pressured. We are "mind on the hoof" (Clark, 1997), and cognition must be understood in terms of how it functions under the pressures of real-time interaction with the environment.
3. We off-load cognitive work onto the environment. Because of limits on our information-processing abilities (e.g. limits on attention and working memory), we exploit the environment to reduce the cognitive workload. We make the environment hold or even manipulate information for us, and we harvest that information only on a need-to-know basis.
4. The environment is part of the cognitive system. The information flow between mind and world is so dense and continuous that, for scientists studying the nature of cognitive activity, the mind alone is not a meaningful unit of analysis.
5. Cognition is for action. The function of the mind is to guide action, and cognitive mechanisms such as perception and memory must be understood in terms of their ultimate contribution to situation-appropriate behavior.
6. Off-line cognition is body based. Even when decoupled from the environment, the activity of the mind is grounded in mechanisms that evolved for interaction with the environment—that is, mechanisms of sensory processing and motor control.

Claims 1 and 6 are perhaps the most significant for understanding representations, as well as the relation of PL to high-level cognitive tasks. We do not attempt any comprehensive analysis here, but limit ourselves to extending the important points derived earlier for understanding PL and cognition.

If the idea of embodied cognition is taken to mean that we do not have any abstract representations, able to be processed separately from the execution of actions, it is probably incorrect, and it would fail to allow a reasonable account of PL effects in high-level domains, for much the same reasons that plague the PSS and classical accounts. Specifically, the selective, abstract, and amodal properties of perceptual representations—the same ones that make the products of perception and PL most useful for complex cognitive tasks—preclude too close a coupling of PL with specific actions. As we will see below, evidence from PL interventions in high-level cognitive domains suggests that when learners come to apprehend important structures, this learning may improve their performance on a variety of tasks, including remote transfer tasks. Structure may be learned and used apart from specific actions. PL phenomena of this sort remind us of the classic work in animal learning indicating that stored representations obtained from perception can be used flexibly for different actions (Tolman, 1948), and, if we can add an update, for thinking. Binding perceptual representations too closely to specific actions would be problematic for reasons analogous to PSS ideas, where rerunning segments of prior perceiving episodes, complete with sensory activations, would seem to impede the extraction of abstract invariance detectable in new contexts. Just as Tolman argued for representations that could not be explained as stimulus–response pairings, embodiment consisting of a necessary connection between perceptual representations and specific actions would fail to provide a reasonable account of perception or PL. That said, many versions of embodied cognition, including most of the claims above, do not mandate such an extreme connection. Indeed, the general idea that advances in understanding may emerge from considering connections among perception, action, and thought is an idea with which we sympathize. For example, our argument regarding the use of spatial representations in symbolic domains such as mathematics might be considered to be related to several of the six claims considered by Wilson (2002).

4.7. Implications for Perceptual Learning

The classical view came with its own view of PL, because all of perception, as opposed to registration of raw sensations, was, in fact, associative learning. This view has been superseded by a generation of direct evidence about

perceptual development, indicating that perceptual systems deliver ecologically meaningful descriptions, even from birth (Bushnell, Sai, & Mullin, 1989; Held, 1985; Kellman & Spelke, 1983; Meltzoff & Moore, 1977; Slater, Mattock, & Brown, 1990; Walk & Gibson, 1961; for a review, see Kellman & Arterberry, 1998). The classic perceptual learning burden of constructing meaningful reality from associating sensations is obviated by an improved picture of early perception.

The revised view of perception as sensitive to information about important environmental properties comes with its own mandate for PL, however. An observer at any time is surrounded by a wealth of meaningful information about objects, surfaces, and events. There are an unlimited number of environmental features and relations that could be important for different tasks. Processes of learning serve to optimize performance of particular tasks by discovering which information is relevant to them, refining and attuning perceptual mechanisms to selectively extract this information, and automating that extraction (E. Gibson, 1969; Kellman & Garrigan, 2009). This kind of PL—that makes perceivers better at discovering and extracting currently available information—is the prevailing notion of PL in contemporary research.

Taken together, contemporary views of perception and PL provide clear foundations for beginning to understand and explore the role of PL in high-level cognitive tasks. The properties of perception that figure prominently are these: Perceptual representations are amodal, abstract, and selective. These are the properties that allow them to be functionally useful in thought and action. Extraction of complex relations connects directly to high-level thinking and underwrites action. If perceptual representations were not in a form that connects to capacities to reason, imagine, and plan, it would be hard to see their point. The synergistic relationship between extraction of important structure and thinking propels learning and the development of expertise.



5. PERCEPTUAL LEARNING AND INSTRUCTION

Mere mention of the word “instruction” evokes an image of teacher speaking to a class. Our ordinary intuitions about teaching and learning in formal settings, as well as most learning research, appear to be colored by a stereotype about what learning is and how it works. Bereiter and Scardamalia (1998) described this stereotype as a “folk psychology” view of learning, specifically, what they termed the “container metaphor”:

Knowledge is most readily conceived of as specifiable objects in the mind, such as discrete facts, beliefs, ideas... (Learning) ... involves retaining and retrieving such objects. (Bereiter & Scardamalia, 1998, p 487).

As we have seen, PL encompasses much that falls outside of this view of learning. Bereiter and Scardamalia contrasted with the conventional “mind as container” view a different idea: “mind as pattern recognizer.” PL is the type of learning that leads to mind as pattern recognizer.

That changes in the way information is extracted are important to expertise has been frequently documented. De Groot (1965), himself a chess master, studied chess players, with the expectation that master level players considered more possible moves and countermoves or in some sense thought more deeply about strategy. Instead, he found that their superiority was shown primarily on the perceptual side. Masters had become able to extract meaningful patterns in larger chunks, with greater speed and less effort than less skilled players. De Groot (1965) suggested that this profile is a hallmark of human expertise in many domains:

We know that increasing experience and knowledge in a specific field (chess, for instance) has the effect that things (properties, etc.) which, at earlier stages, had to be abstracted, or even inferred are apt to be immediately perceived at later stages. To a rather large extent, abstraction is replaced by perception, but we do not know much about how this works, nor where the borderline lies. (pp. 33–34)

Similar differences between experts and novices have since been found in research on expertise in a variety of domains, such as science problem solving (Chi, Feltovich, & Glaser, 1981; Simon, 2001), radiology (Kundel & Nodine, 1975; Lesgold, Rubinson, Feltovich, Glaser, & Klopfer, 1988), electronics (Egan & Schwartz, 1979), and mathematics (Robinson & Hayes, 1978). An influential review of learning and its relation to education (Bransford, Brown, & Cocking, 1999) summed it up this way:

Experts are not simply “general problem solvers” who have learned a set of strategies that operate across all domains. The fact that experts are more likely than novices to recognize meaningful patterns of information applies in all domains, whether chess, electronics, mathematics, or classroom teaching. In De Groot’s (1965) words, a “given” problem situation is not really a given. Because of their ability to see patterns of meaningful information, experts begin problem solving at “a higher place” (DeGroot, 1965). (p. 48)

5.1. Natural Kind Learning

It is interesting that school learning centers so heavily on explicit verbal instruction about facts and procedures, given that more implicit PL

processes appear to dominate the prodigious learning accomplishments of children in the years before they reach school age. Much of early learning may be characterized as discovery processes in PL, and these include apprehension of highly abstract relations, even very early on (Marcus, Vijayan, Bandi Rao & Vishton, 1999). Natural kind learning exemplifies some of the most interesting and powerful characteristics of this kind of learning and transfer. Imagine a young child going for a walk with her father. Upon seeing a dog, the child points, and her father says “That’s a dog.” Suppose this particular dog is a small white poodle. On some other day, the child sees another dog—this one a large black Labrador retriever. Again, someone says “dog.” And so on. With each instance, something about a particular dog (along with the label “dog”) is encoded. As the process continues, and a number of instances (probably not a particularly large number) have been encountered, the child becomes able to look at a new, never before seen dog and say “dog.” This is the magical part, as each new dog will differ in various ways from *any* of the examples previously encountered. Moreover, the child is concurrently coming to distinguish correctly novel instances of dog, cat, squirrel, etc., from each other. A particular cat or squirrel may have properties that resemble some known dog; a small black dog and a large black cat are more similar in color and size than are a large black and small white dog. Despite similarities of instances across different classes and differences among instances within classes, the learner comes to extract properties sufficient to classify novel instances accurately. Much of the relevant PL would seem to require discovery of abstract relations, as simple features, such as color, are seldom the crucial determinants. Shape variables are often important, such as the differing jaw or body structures of dogs and cats. Shape variables are highly relational and abstract, rather than tied to particular colors, sizes, and contexts, which is what allows those who have undergone this kind of learning to effortlessly recognize a glass tabletop ornament as a dog versus a cat.

The properties underlying a classification can be complex and implicit. If a child, or even an adult, is asked to state a set of rules that would allow a novice to distinguish dogs, cats, and wolves, they cannot ordinarily do so. Even the hypothesis about jaw and body structure of dogs and cats, mentioned in the previous paragraph, is a conjecture the authors have generated from poring over examples. For adults, even learning researchers, knowing cat versus dog when one sees them is easy, but furnishing an account in declarative knowledge or a diagnostic procedure is hard, and it is not a typical accompaniment of the ability to recognize.

Nor do the toddler's striking feats of natural learning occur from being given lectures on the distinguishing features of dogs or cats. Rather, structure is extracted from encountering instances and receiving category feedback. Such PL processes are crucial not only for developing understanding of the objects and events in the world; they also play a pivotal role in language acquisition, at multiple levels. Concepts like noun, verb, adverb, and preposition are taxing enough when taught explicitly in middle school. How is it that these abstract classes are extracted and used in language acquisition, allowing grammatical structures to be processed (e.g. Hirsh-Pasek & Golinkoff, 1997) and facilitating the learning of new words? At a different level, learning may be involved in the ability of the young language learner to detect invariance in the structure of speech signals across different speakers. Evidence suggests that the PL processes needed for these achievements, including clear cases of abstract PL, are present relatively early in infancy (Gomez & Gerken, 1999; Marcus et al., 1999; Saffran, Loman, & Robertson, 2000).

5.2. Relations among Types of Learning: Toward a “Fundamental Theorem of Learning”

When a child starts school or other formal learning, the focus of most instructional efforts, as it has been in most research on instruction, is on declarative and procedural activities. This emphasis can be seen, in part, as fitting with important patterns that scientists have discovered regarding human cognitive development. Before a certain age, the introduction of formal concepts and reasoning is likely to be pointless (NRC, 2001; Piaget, 1954).

Conversely, PL is among types of learning that seem to operate from the beginning of life, and it plays an important role in natural kind learning, language acquisition, and transactions with many kinds of objects and events. When a child has reached school age, it might be assumed that with those foundations already in place, “higher” cognitive activities—encompassing explicit facts, concepts, procedures, and thinking—take center stage.

We believe it would be a misunderstanding, however, to believe that when more explicit aspects of learning are introduced, the PL components of learning fade into the background. Although we do not attribute this view explicitly to anyone, it may be natural to assume that by school age, perceptual transactions with the environment have been largely mastered or that they operate in a relatively steady-state fashion. A related point may

be made about the content of thought and learning. Theories of cognitive development have tended to be saturated with classical views of perception (Kellman & Arterberry, 1998); thus, in Piaget's views, and subsequent related views, the early role of the senses is in associative "sensorimotor" transactions. Conceptual inputs in various domains must operate to make perceptual data useful for abstract knowledge (e.g. Leslie, 1995; Mandler, 1988, 1992; Piaget, 1952, 1954). Research that has produced a radically different understanding of the perceptual starting points of development changes this picture and has profound implications for cognitive development, which have been discussed elsewhere (e.g. Jones & Smith, 1993; Kellman & Arterberry, 1998). In the present context, the crucial consequences of the contemporary understanding of perception as delivering abstract structural knowledge are that 1) the perceptual part of learning remains important in most or all learning domains, and 2) the products of perception are not static or previously mastered, but are dynamically changing as an important part of learning in any domain.

Perhaps most interesting and important, the changes in perceptual pickup and the use of declarative and procedural knowledge and reasoning should not be considered unrelated aspects of task performance. There is a crucial, interactive relationship between these, one that parallels the close coupling of perception and action (J. Gibson, 1966, 1979). Although it has seldom been emphasized in learning research, PL processes—that attune the encoding, classification, discrimination, or recognition of incoming information—bear a pervasive relationship to the better-known declarative and procedural aspects of learning. Only half jokingly, we call this the "Fundamental Theorem of Learning." It states that

All effective use of declarative and procedural learning presupposes pattern recognition.

Suppose a learner in some domain has acquired a vast array of facts, concepts, and procedures. How are these deployed? How do they lead to effective problem solving in new situations as they arise? Randomly producing facts and procedures is at best inefficient and at worst pathological. Obviously, facts and procedures must be used selectively and appropriately. Accomplishing appropriate selection depends on accurate classification of problems or situations. When one is confronted with a new problem or situation, which facts apply? Which procedures are relevant? Fundamentally, these are questions of encoding and classifying the input; they require recognizing, amidst irrelevant detail, the structural patterns that matter. They are pattern recognition problems.

Becoming able to see what matters in a given situation has long been regarded as the essence of meaningful learning and creative problem solving (c.f. [Duncker, 1945](#); [Wertheimer, 1959](#)). What has often been missing from the discussions of the role of seeing in problem solving is the learning process that allows the learners to become able to recognize, in new situations, the meaningful structures that matter and to distinguish the relevant from the irrelevant. This is the role of PL, and our statement of this “fundamental theorem” is simply a reminder that even in high-level learning tasks and domains, processes that advance encoding, discrimination, classification, and structure recognition allow facts, concepts and procedures to be used effectively.

5.3. Perceptual Learning Technology

Modeling PL is a complicated and unfinished effort ([Ahissar & Hochstein, 2004](#); [Fahle & Poggio, 2002](#); [Kellman & Garrigan, 2009](#); [Petrov et al., 2005](#)). This is especially true for perceptual classifications that are based on abstract relations ([Kellman, Burke, & Hummel, 1999](#); for discussion, see [Kellman & Garrigan, 2009](#)). There are currently relatively few models that even purport to discover abstract relationships that govern a classification, even in restricted domains. Improving our understanding of such abilities will be valuable for many scientific and technological reasons. For example, in computer vision and artificial intelligence, we still lack systems that can learn to recognize cats in ordinary scenes, much less learn to classify a glass table ornament as a cat, and we are far away from being able to extract even more abstract regularities, such as when a tone of voice conveys sarcasm.

Fortunately, the task of understanding the conditions under which PL occurs and the variables that affect it is a much more tractable one than developing models of high-level PL. Understanding the principles of PL is an active area of research (e.g. [Ahissar & Hochstein, 2004](#); [Mettler & Kellman, 2010](#); [Seitz & Watanabe, 2005](#); [Zhang et al., 2010](#)).

Some efforts have focused on complex, real-world tasks, attempting to systematically address PL and accelerate the growth of perceptual expertise in instructional settings. These efforts have already produced remarkably successful outcomes in a variety of learning domains.

[Kellman and Kaiser \(1994\)](#) developed PLMs to address difficult problems in aviation training. In a Visual Navigation PLM, pilots learned navigational skills by mapping, on short, speeded trials, videotaped segments of out-of-the-cockpit views onto locations shown on standard visual navigation (Visual Flight Rules sectional) charts. Remarkable improvements in

accuracy and speed occurred in less than an hour of training, even among experienced aviators. In an Instrument Flight PLM, the focus was on flight instrument interpretation. On short speeded trials, pilots classified aircraft attitude (e.g. climbing, turning) from an array of primary flight displays used by pilots to fly in instrument conditions. They found that under an hour of training allowed novices to process configurations more quickly than and just as accurately as civil aviators who had on average 1000 h of flight time (but who had not used the PLM). When experienced pilots used the PLM, they also showed substantial gains, paring 60% off the time needed to interpret instrument configurations.

PL interventions to address speech and language difficulties have been shown to produce benefits (Merzenich et al., 1996; Tallal, Merzenich, Miller, & Jenkins, 1998). For example, Tallal et al. showed that auditory discrimination training in language learning using specially enhanced and extended speech signals improved both auditory discrimination performance and speech comprehension in language-impaired children.

Applications in medical and surgical training illustrate the value of PL in addressing dimensions of learning not encompassed by ordinary instruction. Guerlain et al. (2004) applied PLM concepts to address issues of anatomic recognition in laparoscopic procedures. They found that a computer-based PLM approach patterned after the work of Kellman and Kaiser (1994) produced better performance than traditional approaches. The training group presented with variation in instances selected to encourage learning of underlying invariance later showed improvement on perceptual and procedural measures, whereas a control group who saw similar displays but without the structured PLM did not. Their data implicated PL as the source of the improvement, as neither group advanced on strategic or declarative knowledge tests.

More recently, Krasne et al. (submitted) developed and tested two computer-based perceptual/adaptive learning modules (PALMs) in the preclerkship curriculum for all first- and second-year medical students at the UCLA School of Medicine. One module focused on pathologic processes in skin histology images (*Histopathology PALM*) and the other for identifying skin-lesion morphologies (*Dermatology PALM*). The goal was to assess students' ability to develop pattern recognition and discrimination skills leading to accuracy and fluency in diagnosing new instances of disease-related patterns. These were short learning interventions, with objective learning criteria typically achieved in 15–35 min. Results indicated strong learning gains in accurately classifying previously unseen cases, elevating students' performance

in both the first and second years of medical school well beyond the levels attained from conventional instruction alone. There were strong gains in both accuracy and fluency; besides becoming more accurate, learners averaged a 53% reduction in classification time across both years and PALMs. Effect sizes averaged in the 1.0–1.5 range for both accuracy and fluency. These results with brief interventions suggest that PL interventions impact aspects of learning that are not well addressed by conventional instruction. They also suggest remarkable promise for the use of PL to improve learning in a variety of medical and other domains.

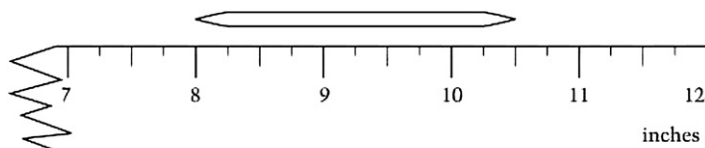
Over the past decade, we have undertaken large-scale, systematic efforts to study and apply PL technology in mathematics and science learning (Kellman et al., 2009; Massey et al., 2011; Silva & Kellman, 1999; Wise et al., 2000). Although these subjects involve a variety of cognitive processes, they rely substantially on pattern recognition and fluent processing of structure, as well as mapping across transformations (e.g. in algebra) and across multiple representations (e.g. graphs and equations). Few instructional activities directly address these aspects of learning, and a variety of indicators suggest that they may be disproportionately responsible for students' difficulties in learning (Kellman et al., 2009). Findings consistently indicate that even short PLM interventions can accelerate fluent use of structure in contexts such as the mapping between graphs and equations (Kellman et al., 2008; Silva & Kellman, 1999), apprehending molecular structure in chemistry (Russell & Kellman, 1998; Wise et al., 2000), processing algebraic transformations, and understanding fractions and proportional reasoning (Kellman et al., 2008, 2009; Massey et al., 2011). Earlier, we presented the example of an *Algebraic Transformations PLM*. To convey the scope and approach of PLMs in mathematics learning, we describe here one other PLM in detail and summarize some others. These examples will help to illustrate both the learning effects from PLMs as well as their distinctive features as learning interventions.

An illuminating example is a PLM that we developed to help elementary students master linear measurement with rulers of varying scales. When one considers a standard ruler, it is a rather remarkable device that organizes a numerical symbol system in a spatial layout—essentially the positive side of a rational number line on a strip of wood or plastic. The continuous extent is evenly partitioned into units and marked by numbered hash marks, with hash marks of several sizes arranged to indicate different scales layered on the same ruler (e.g. half inches, quarter inches, eighth inches). Once one has acquired expertise with this instrument, it is a simple matter to “just

see” the structure. The inches or quarter inches or centimeters or meters are readily perceived as objects that can be manipulated and enumerated in various ways to measure linear extents.

As countless teachers can testify, however, acquiring such insight is not a simple or reliable achievement for many elementary or even middle school students, despite conscientious instruction. An indication of the learning difficulty comes from results on the National Assessment of Education Progress (<http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=mathematics>), on a released item known as the broken ruler problem. A version of the problem is illustrated in Figure 4.3. A toothpick is pictured above a standard 12-inch ruler that has been broken so that the left-hand edge starts at 7 inches. The toothpick is positioned so that it starts at 8 and ends at $10\frac{1}{2}$, and students are asked to enter the length of the ruler. Alarming, only 20% of 4th graders and 58% of 8th graders give the correct answer. Of particular interest are the two most common incorrect answers: $10\frac{1}{2}$ and $3\frac{1}{2}$. Students who give the former answer are most likely following a poorly understood, inflexible procedure that involves reading the rightmost endpoint as the length—simply ignoring that the ruler is broken. Students who say that the toothpick is $3\frac{1}{2}$ inches long are probably relying on a counting routine and counting the hash marks starting with the leftmost hash mark as “1.” (It is a common classroom observation that students are extremely puzzled as to why the left-most edge of a ruler is 0 rather than 1, and why they are told to line things up starting at 0. After all, when counting discrete items, one always starts with one, not zero.)

Both of these incorrect answers indicate that the students are not perceiving units on the ruler that have extent. From conventional instruction, they have picked up some aspects of measurement facts and procedures, but the mapping of what they have learned onto structure in the problem is faulty. They do not recognize that an inch (or centimeter, etc.) on a ruler is the extent between the hash marks that demarcate the unit, not just the point where the numbered hash mark is located. The beauty of the broken ruler problem is that it reveals this; students succeed with much higher accuracy if they are given an ordinary ruler problem, in which the zero point lines up with the left edge of the toothpick. A related and persistent difficulty is that students struggle to map fractions to rulers. Difficulties with fraction notation aside, if a student does not see an extended unit to begin with, he or she will have difficulty identifying the subpartitions of units that map to fractional quantities.



What is the length of the toothpick in the figure above?

Figure 4.3 The “broken ruler” problem. Released measurement item from the 2003 National Assessment for Educational Progress (US Department of Education, IES, National Center for Education Statistics, <http://nces.ed.gov/nationsreportcard/itmrlsx>).

To address this problem of seeing the relevant structure, we developed learning software that presents students with many short, interactive, animated learning trials in which students interact with the key structures and relationships underlying linear measurement. A typical trial presents the student with a graphic display showing a ball on top of a ruler and billiard cue poised to strike it. The student is given either a starting point and an ending point and asked to say the distance traveled, or they are given a starting point and a traveling distance and are asked to say what the endpoint will be. The learning items in the database vary with the types of values involved, whether the rulers are fully versus partially labeled, and whether they are partitioned in the most economical way to solve the problem or are overpartitioned (e.g. a ruler marked in units of sixteenths for a problem involving eighths). Movement on the ruler can be either rightward or leftward. The quantities involved vary from single digits into the hundreds and included both fractions and integers. Learners receive immediate animated feedback on each trial and are also given periodic feedback on their progress through the module.

Instead of emphasizing verbal explanations or procedural calculations, this *Linear Measurement PLM* concentrates the students’ attention and effort on learning to pick up relevant structures and relationships. The items in the learning set are designed so that each student sees many varied examples; these are conditions in which PL processes come to discover and fluently extract important structures in different contexts. In this way, PLMs accelerate students’ expertise until they are able to “just see” what is important and relevant in each problem.

In a formal study, 63 sixth-grade students in a low performing urban middle school completed a pretest and used the *Linear Measurement PLM*,

then took an immediate post-test as well as a delayed post-test a full 4 months later. The 6th graders were compared with a group of forty-nine 7th and 8th graders in the same school who took the assessment without using the module. The assessment, which included many transfer items that did not resemble the learning trials, tested children's ability to use a partitioned number line to express the length of a line segment in generic units; to use both conventional and broken rulers to measure lengths in inches and centimeters; to use conventional and broken rulers to construct extents with varying lengths; to solve addition and subtraction problems with fractions; and to solve open-ended word problems involving linear measurements. Both the 6th grade intervention students at pretest and the older control students scored <50% on the assessment. After completing the module, the 6th graders' scores improved dramatically (Figure 4.4), with effect sizes (Cohen's d) comparing pretest scores versus post-test scores and intervention versus control groups ranging from 0.86 to 1.06 (Kellman, Massey & Son, 2009; Massey, Kellman, Roth, & Burke, 2011). The studies also demonstrated remarkable durability of learning: Scores on delayed post-tests conducted 4 months later, with no intervening study activities, indicated that the learning gains for the intervention groups were fully maintained.

Other PLM interventions in mathematics learning have produced comparable results. In the area of fractions and measurement, PLMs focusing on partitioning and iterating units and mapping equivalent quantities across different units not only produced effect sizes in the range of 1.0 to 2.8 but led to remote transfer of learning to multiplying and dividing fractions and mixed numbers (tasks that were not part of the PLM). Moreover, as in the case of the *Linear Measurement PLM* described above, the learning gains showed no decrements in delayed post-tests administered 4–5 months later. Both the remote transfer and durability of the learning highlight important characteristics of PL: Becoming able to see relevant structure in a domain allows that structure to be used in varied tasks and comprises an enduring kind of learning.

5.4. Elements of Perceptual Learning in Instruction

These examples of PLMs in real-world learning contexts illustrate some of the conditions that produce PL effects and some of the characteristics of learning attainments from these interventions. More generally, what are the elements of PL interventions?

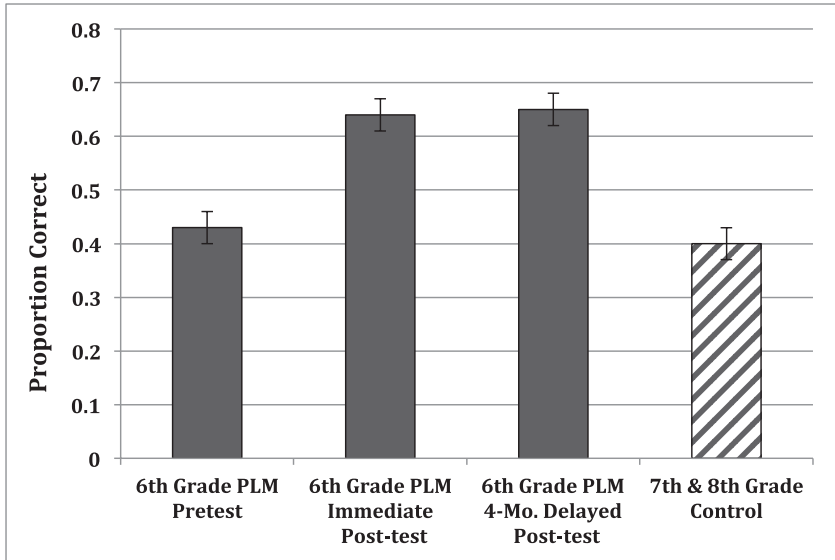


Figure 4.4 Results from a Study of the Linear Measurement PLM. Pretest, immediate post-test, and delayed post-test scores are shown for the 6th grade intervention group compared to 7th and 8th grade students in the same school. The delayed post-test was administered after a delay of 4 months, with no intervening study activities. Error bars indicate ± 1 standard error of the mean (Adapted from Kellman, Massey & Son, *TopicCS in Cognitive Science*, 2009; Cognitive Science Society, Inc., p. 16).

Kellman et al. (2009) argued that at least three general properties are crucial.³ The most basic requirement is that PL tasks focus on the extraction of structure. PLMs involve encoding, discrimination, comparison, and/or classification. A contrast in mathematics learning is that PL interventions need not involve computation of numerical answers. In PL tasks, the learner engages in practice with displays or representations in which success depends on the learner coming to attend to, discriminate, classify, or map structure. Utilizing structure is of course involved in other types of instruction, but

³This section focuses on characteristics that define PL interventions as a distinctive type of learning activity. Many more specific features of PLMs, not discussed in detail here, serve to optimize learning in this general format and to configure them for particular learning challenges and goals. These include issues of sequencing, feedback, variation of positive and negative instances of categories, mixing of learning tasks, integration of PL activities with conventional instruction, and so on. A number of features of PL technology and related adaptive learning technology are covered by US patent #7052277 and patents pending, assigned to Insight Learning Technology, Inc. For information, please contact the authors or Info@insightlearningtech.com.

a PL task focuses on commonalities and variations in structure as its primary learning content. A second characteristic is that PLMs tend to involve numerous short classification trials with varied instances. The learner makes classifications on these trials and (in most cases) receives feedback. Systematic variation across learning instances is crucial, because in most real-world tasks, PL involves the discovery of invariance amidst variation (Gibson, 1969). Discovery processes require sufficient variation for relevant properties to be disentangled from irrelevant ones. This aspect of PL interventions is most powerful in producing transfer of learning to new situations that involve common or related structures. Emphasis on the discovery of invariance amidst robust variation is crucial in realistic learning tasks, but it differs from most contemporary laboratory studies of PL, which typically target simple sensory discriminations and involve large numbers of trials with a small set of fixed stimuli (e.g. Fahle & Poggio, 2002; for a discussion, see Garrigan & Kellman, 2008). Finally, PL interventions tend to have minimal emphasis on explicit instruction. The learning comes from transactions with the input, not verbal exchanges. The primary task in a PL intervention does not involve verbal or written explanations of facts, concepts, or procedures. This is a major difference from conventional instruction, which is dominated by explicit description (which also addresses important aspects of learning). PL interventions may incorporate explicit introductions or brief discussions, but these do not comprise the central learning tasks nor are they capable of producing the results obtained with PLMs.

Another important question is, “How do we know that PL effects are occurring from a learning intervention?” In complex tasks and realistic learning settings, we have less control over materials and activities than in most laboratory situations. Moreover, we would expect, as we have argued in this chapter, that PL works synergistically with other processes of learning and thinking in domains such as mathematics. Given these background conditions, it is unlikely that any intervention in a complex-learning domain targets PL uniquely, and it is difficult to claim that any learning gains are solely the result of PL. Likewise, although the issue has not received much attention, it would be hard to claim that effects produced by other instructional interventions do not involve a PL component. PLMs attempt to condense or accelerate PL, but PL no doubt goes on less systematically in other learning situations.

Synergies aside, there appear to be some characteristic signatures of PL effects. Kellman et al. (2009) suggested four of these, summarized in Table 4.2.: 1. *Generativity in structure use*. PLMs in complex learning

domains are designed to improve pickup and processing of structural invariants across variable contexts. A hallmark of successful PL is the evidence of accurate and/or fluent classification of novel cases. Moreover, PLMs often facilitate remote transfer to different-looking problem types that involve the same underlying structure. Such transfer is a notorious problem in settings using conventional declarative instructional approaches. Evidence of accurate and fluent classification of novel instances, and transfer to contexts involving different procedural requirements but common structures, provides evidence of PL. 2. *Fluency effects.* PL effects typically include improved fluency of information extraction (indicated in measures of speed, parallel processing, or reduced effort or cognitive load). Acquisition data within PLMs suggest that fluency in information extraction increases gradually across interactive trials. Gradual improvement is not unique to PL but does contrast with some effects of declarative instruction, in which a learner may either know or not know a certain concept. A particularly clear case of PL effects on fluency can be made when relevant declarative knowledge is already present prior to an intervention, and a PL intervention produces improved fluency, as in the *Algebraic Transformations PLM* described

Table 4.2 Some Possible Signature Effects of Perceptual Learning Interventions. The effects shown are common outcomes of PL interventions that tend to distinguish them from outcomes of instruction focused on declarative or procedural knowledge (see text).

Generativity in use of structure

- Accurate and/or fluent processing of novel cases
 - Improvement on unpracticed tasks that involve learned structures
-

Improvements in fluency

- Faster processing
 - Greater parallel processing
 - Reduced cognitive load or effort
-

Implicit pattern recognition versus explicit knowledge

- Improved performance without new explicit declarative or procedural knowledge
-

Durability of learning

- Improved information extraction and structural intuition that persist over long delays and are highly resistant to forgetting
-

earlier. 3. *Implicit pattern recognition versus explicit knowledge.* Although PL may provide important scaffolding for explicit, verbalizable knowledge, PL itself need not involve changes in explicit knowledge. PL changes the way a learner views a problem or representation, and these changes need not be accompanied by new explicit facts, concepts, or procedures. Transfer tests routinely indicate this dissociation from PL interventions (Guerlain et al., 2004; Kellman et al., 2009). 4. *Delayed testing effects.* Common wisdom has it that one never forgets how to ride a bicycle. If true, riding a bicycle, a task that clearly involves considerable PL, differs from most declarative and procedural learning. It is not by accident that math teachers spend the first month of a new school year reviewing content from the prior year. Facts and procedures are subject to substantial forgetting over a period as long as a summer vacation from school. Although more research is needed, there are indications that the improved facility in picking up patterns and structure from PL, like riding a bicycle, may be less subject to decay over time. In the measurement and fraction PLMs described above, we have consistently observed no decrements in learning gains when students were tested after 4- to 5-month delays (Kellman et al., 2009; Massey et al., 2011).

It is also possible to test directly for PL effects. In domains where the central task is clearly focused on classification, such as the *Dermatology and Histopathology PALMs* described above, rapid and accurate classification of new instances illustrates straightforwardly that learners have improved in the pickup of information. For PL interventions in cognitive domains that also involve symbolic material and substantial reasoning components, the situation is more complicated in attributing learning gains to specifically PL effects. In applying PL technology to such domains, investigators have usually had as a first priority showing that PL leads to meaningful learning gains, beyond those of conventional instruction. Thus, tests of learning and transfer have typically assessed learning on important domain-relevant tasks; in mathematics PLMs, these have involved tasks such as solving algebra problems, performing operations with fractions or measurement, or generating a correct graph from an equation or an equation from a word problem (Kellman et al., 2008). However, more basic psychophysical tests in complex PL domains are also possible. Thai, Mettler, and Kellman (2011) showed that PL interventions, like those we have used in complex, symbolic domains, produce basic changes in information extraction. Participants were trained to classify Chinese characters, based on either overall configurations (structures), featural relations (components),

or nonrelational information (stroke count), used as a control. PLM participants showed strong domain-relevant learning gains in discriminating and classifying Chinese characters. Before and after training, however, they were also tested for basic changes in information extraction using a visual search task, which had not been part of training. Search displays contained all novel exemplars, involved manipulations of target-distractor similarity using structures and components, and included heterogeneous and homogeneous distractors. Robust improvements in visual search for structure and component PL training were found relative to a control group that did not undergo PLM training. These results provide direct evidence that high-level PL interventions improve learning by altering extraction of information, including changing perceptual sensitivity to important relational structures. This study is interesting in connecting a high-level cognitive task to changes in information pickup detectable by more basic psychophysical methods. Further research of this sort may prove useful, both in understanding the synergies of various cognitive abilities and in optimizing PL interventions.

Another significant issue for further research is how PL interventions might best be combined with other modes of instruction. Acquiring declarative and procedural knowledge, improving critical thinking, and other aspects of learning do not become less important because we are coming to understand that neglected PL components are crucial in many learning domains. In fact, it seems likely that instructional methods of all types will benefit from understanding more clearly these different components of learning and their interactions. A discussion of explicit concepts, or a proof, may be easier when the teacher knows that the student is correctly mapping the words onto problem structure, and a procedure may be better understood, better remembered, and certainly better applied, when the student can see where and why it applies.



6. CONCLUSION

Research in PL offers previously unsuspected synergies with high-level cognitive tasks and processes. Through an emerging technology of PL, it also promises remarkable potential to improve learning in almost any domain, including complex, symbolic ones. To understand and utilize these possibilities fully requires making clear basic connections between perception, cognition, and learning, especially the implications of contemporary views of perception as abstract, amodal, and selective. In this chapter, we

have tried to describe these connections in ways that allow us to integrate and illuminate recent research and applications of PL. We hope these efforts contribute to future progress in understanding cognition, perception, and learning.

ACKNOWLEDGMENT

The authors gratefully acknowledge support from: the US Department of Education, Institute of Education Sciences (IES), Cognition and Student Learning Program awards R305A120288 and R305H060070; the National Science Foundation (NSF) Research on Education and Evaluation in Science and Engineering (REESE) Program award 1109228; the National Institute of Child Health and Human Development (NICHD) award 5RC1HD063338; the UCLA Center for Advanced Surgical and Interventional Technology (CASIT); and the Institute for Innovative Technology in Medical Education (iInTIME); as well as the US Department of Education, IES SBIR award ED-IES-10-C-0024 to Insight Learning Technology, Inc. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the US Department of Education, the NSF or other agencies. We thank Patrick Garrigan, Keith Holyoak, Hongjing Lu, K.P. Thai, Everett Mettler, and Gennady Erlikhman for helpful discussions, and Rachel Older for general assistance.

REFERENCES

- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10), 457–464.
- Ahissar, M., Laiwand, R., Kozminsky, G., & Hochstein, S. (1998). Learning pop-out detection: building representations for conflicting target-distractor relationships. *Vision Research*, 38(20), 3095–3107.
- Baron, J. (1978). The word-superiority effect: perceptual learning from reading. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Linguistic functions in cognitive theory* (Vol. 6, pp. 131–166). Hillsdale, NJ: Erlbaum.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–660.
- Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 358(1435), 1177–1187.
- Bereiter, C., & Scardamalia, M. (1998). Rethinking learning. In D. Olson, & N. Torrance (Eds.), *The handbook of education and human development* (pp. 485–512). Malden, MA: Blackwell.
- Berkeley, G. (1709/1910). *Essay towards a new theory of vision*. London: Dutton.
- Bransford, J. D., Brown, D., & Cocking, R. (1999). *How people learn: Mind, brain, experience, and school*. Washington, DC: National Research Council.
- Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: the acquisition of a hierarchy of habits. *Psychological Review*, 6(4), 345–375.
- Bushnell, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, 7(1), 3–15.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152.

- Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, 375, 121–123.
- van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, 32(3), 579–590.
- De Groot, A. D. (1965). *Thought and choice in chess*. Amsterdam: Noord-Hollandsche Uitgeversmaatschappij.
- Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58(5), i-113. (Lees, L. S., Trans.).
- Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition*, 7(2), 149–158.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Fahle, M., & Poggio, T. (Eds.). (2002). *Perceptual learning*. Cambridge, MA: The MIT Press.
- Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, 22(13), 5749–5759.
- Garrigan, P., & Kellman, P. J. (2008). Perceptual learning depends on perceptual constancy. *Proceedings of the National Academy of Sciences of the United States of America*, 105(6), 2248–2253.
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6), 278–285.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Prentice-Hall.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Oxford: Houghton Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Goldstone, R. (2000). Unitization during category learning. *Journal of Experimental Psychology*, 26(1), 86–112.
- Goldstone, R. L., Landy, D., & Son, J. Y. (2008). A well grounded education: the role of perception in science and mathematics. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 327–355). New York: Oxford University Press.
- Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-old leads to specific and abstract knowledge. *Cognition*, 70, 109–135.
- Guerlain, S., La Follette, M., Mersch, T. C., Mitchell, B. A., Poole, G. R., Calland, J. F., et al. (2004). Improving surgical pattern recognition through repetitive viewing of video clips. *IEEE Transactions on Systems, Man, and Cybernetic, Part A: Systems and Humans*, 34(6), 699–707.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243–259.
- He, S., & MacLeod, D. I. A. (2001). Orientation-selective adaptation and tilt after-effect from invisible patterns. *Nature*, 411, 473–476.
- Held, R. (1985). Binocular vision: behavioral and neuronal development. In J. Mehler, & R. Fox (Eds.), *Neonate cognition: beyond the blooming buzzing confusion* (pp. 37–44). Hillsdale, NJ: Erlbaum.
- von Helmholtz, H. (1864/1962). In J. P. C. Southall (Ed.), *Treatise on physiological optics*. (Vol. 3). New York: Dover Publications, Inc.
- Hirsh-Pasek, K., & Golinkoff, R. (1997). *The origins of grammar*. Cambridge, MA: The MIT Press.
- Hochberg, J. E. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception* (pp. 309–331). New York: Holt, Rinehart, and Winston.
- Johansson, G. (1970). On theories of visual space perception. A letter to Gibson. *Scandinavian Journal of Psychology*, 11, 67–74.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14, 201–211.

- Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development, 8*(2), 113–139.
- Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception*. New York: Praeger.
- Keane, B. P., Lu, H., Pappathomas, T. V., Silverstein, S. M., & Kellman, P. J. (2012). Is interpolation cognitively encapsulated? Measuring the effects of belief on Kanizsa shape discrimination and illusory contour formation. *Cognition, 123*(3), 404–418.
- Kellman, P. J. (2002). Perceptual learning. (3rd ed). In H. Pashler, & C. R. Gallistel (Eds.), *Stevens' handbook of experimental psychology* (Vol. 3, pp. 259–299). New York: John Wiley & Sons.
- Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy*. Cambridge, MA: The MIT Press.
- Kellman, P. J., Burke, T., & Hummel, J. (1999). Modeling the discovery of abstract invariants. In B. Stankewicz & T. Sanocki (Eds.), *Proceedings of the 7th annual workshop on object perception and memory (OPAM)* (pp. 48–51).
- Kellman, P. J., & Garrigan, P. B. (2009). Perceptual learning and human expertise. *Physics of Life Reviews, 6*(2), 53–84.
- Kellman, P. J., Garrigan, P., & Shipley, T. F. (2005). Object interpolation in three dimensions. *Psychological Review, 112*(3), 586–609.
- Kellman, P. J., & Kaiser, M. K. (1994). Extracting object motion during observer motion: combining constraints from optic flow and binocular disparity. *Journal of the Optical Society of America, A, Optics, Image Science and Vision, 12*(3), 623–625.
- Kellman, P. J., Massey, C. M., Roth, Z., Burke, T., Zucker, J., Saw, A., et al. (2008). Perceptual learning and the technology of expertise: studies in fraction learning and algebra. *Pragmatics and Cognition, 16*(2), 356–405.
- Kellman, P. J., Massey, C. M., & Son, J. (2009). Perceptual learning modules in mathematics: enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science (Special Issue on Perceptual Learning), 2*(2), 285–305.
- Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23*(2), 141–221.
- Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15*(4), 483–524.
- Kim, D., Imai, H., Sasaki, Y., & Watanabe, T. (2012). Roles of inhibitory processes in perceptual learning. *Journal of Vision, 12*(9), 1128.
- Klatzky, R. L., Wu, B., & Stetten, G. (2008). Spatial representations from perception and cognitive mediation: the case of ultrasound. *Current Directions in Psychological Science, 17*(6), 359–364.
- Knudsen, E. I. (1982). Auditory and visual maps of space in the optic tectum of the owl. *Journal of Neuroscience, 2*(9), 1177–1194.
- Knudsen, E. I., & Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science, 230*(4725), 545–548.
- Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt, Brace and World.
- Krasne, H., Rimoin, B., Kim, D., Craft, N., Kellman, P. J., Drake, T., et al. Perceptual and adaptive learning technology in medical education: the efficacy of pattern recognition in teaching dermatology and histopathology to medical students, submitted.
- Kundel, H., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*, 527–532.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571–579.
- Landau, B. (1999). Reinventing a broken wheel. *Behavioral and Brain Sciences, 22*(4), 623–624.
- Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 720–733.
- Lehar, S. (1999). Gestalt isomorphism and the quantification of spatial perception. *Gestalt Theory, 21*(2), 122–139.

- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., & Klopfer, D. (1988). Expertise in a complex skill: diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Leslie, A. M. (1995). A theory of agency. In A. M. Leslie (Ed.), *Causal cognition: A multidisciplinary debate* (pp. 121–149). New York: Oxford University Press.
- Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision 'template' for position discrimination. *Nature Neuroscience*, 7(2), 178–183.
- Locke, J. (1690/1971). *An essay concerning human understanding*. New York: World Publishing Co.
- Mandler, J. M. (1988). How to build a baby: on the development of an accessible representational system. *Cognitive Development*, 3(2), 113–136.
- Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, 99(4), 587–604.
- Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77–80.
- Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- Massey, C. M., Kellman, P. J., Roth, Z., & Burke, T. (2011). Perceptual learning and adaptive learning technology: developing new approaches to mathematics learning in the classroom. In N. L. Stein (Ed.), *Developmental and learning sciences go to school: Implications for education*. New York: Taylor & Francis.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75–78.
- Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25(6), 954–962.
- Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science*, 271(5245), 77–81.
- Mettler, E., & Kellman, P. J. (2010). Adaptive sequencing in perceptual learning. *Journal of Vision*, 10(7), 1098.
- Michotte, A. (1963). *The perception of causality*. Oxford, England: Basic Books.
- Michotte, A., Thines, G., & Crabbe, G. (1964). Les complements amodaux des structures perceptives. In *Studia psychological*. Louvain, Belgium: Institut de Psychologie de l'Université de Louvain.
- Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision*, 10(1), 51–58.
- Moravec, H. (1988). *Mind children: The future of robot and human intelligence*. Cambridge: Harvard University Press.
- Murray, M., Foxe, D. M., Javitt, D. C., & Foxe, J. J. (2004). Setting boundaries: brain dynamics of modal and amodal illusory shape completion in humans. *The Journal of Neuroscience*, 24(31), 6898–6903.
- Nanay, B. (2010). Perception and imagination: amodal perception as mental imagery. *Philosophical Studies*, 150, 239–254.
- National Research Council. (2001). Adding it up: helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: National Academy Press.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York: W. H. Freeman.
- Ohlsson, S. (1999). Selecting is not abstracting. *Behavioral and Brain Sciences*, 22(4), 630–631.

- Paas, G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133.
- Palmer, E. M., Kellman, P. J., & Shipley, T. F. (2006). A theory of dynamic occluded and illusory object perception. *Journal of Experimental Psychology: General, 135*(4), 513–541.
- Petrov, A. A., Doshier, B. A., & Lu, Z. L. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychological Review, 112*(4), 715–743.
- Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Pizlo, Z. (2010). *3D shape: Its unique place in visual perception*. Boston, MA: The MIT Press.
- Pouget, A., Deneve, S., & Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience, 3*, 741–747.
- Prinz, J. J. (2004). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: The MIT Press.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: a functionalist interpretation. *Cognition, 6*(3), 189–221.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*(2), 275–280.
- Ringach, D. R., Hawken, M. J., & Shapley, R. (2003). Dynamics of orientation tuning in macaque V1: the role of global and tuned suppression. *Journal of Neurophysiology, 90*, 342–352.
- Robinson, C. S., & Hayes, J. R. (1978). Making inferences about relevance in understanding problems. In R. Revlin, & R. E. Mayer (Eds.), *Human reasoning*. Washington, DC: Winston.
- Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance, 7*(4), 733–740.
- Russell, A. A., & Kellman, P. J. (1998, March). Teaching automatic, 3-D recognition of chemical structures, Presented at the American Chemical Society Meeting, Dallas, Texas.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, New Series, 274*(5294), 1926–1928.
- Saffran, J. R., Loman, M. M., & Robertson, R. R. W. (2000). Infant memory for musical experiences. *Cognition, 77*(1), 15–23.
- Sanada, T. M., & Ohzawa, I. (2006). Encoding of three-dimensional surface slant in cat visual areas 17 and 18. *Journal of Neurophysiology, 95*(5), 2768–2786.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1–66.
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*(8), 209–309.
- Seitz, A., & Watanabe, T. (2005). A unified model of task-irrelevant and task-relevant perceptual learning. *Trends in Cognitive Science, 9*(7), 329–334.
- Shepard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91*(4), 417–447.
- Silva, A. B., & Kellman, P. J. (1999). Perceptual learning in mathematics: the algebra-geometry connection. In M. Hahn, & S. C. Stoness (Eds.), *Proceedings of the 21st annual conference of the cognitive science society* (pp. 683–688). Mahwah, NJ: Lawrence Erlbaum Associates.
- Simon, H. A. (2001). Observations on the sciences of science learning. *Journal of Applied Developmental Psychology, 21*(1), 115–121.
- Slater, A., Mattock, A., & Brown, E. (1990). Size constancy at birth: newborn infants' responses to retinal and real size. *Journal of Experimental Child Psychology, 49*(2), 314–322.

- Spelke, E. S. (1987). The development of intermodal perception. In L. B. Cohen, & P. Salapatek (Eds.), *Handbook of infant perception*. New York: Academic Press.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge: MIT Press.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–266.
- Stoffregen, T. A., & Bardy, B. G. (2001). On specification and the senses. *Behavioral and Brain Sciences*, *24*, 195–261.
- Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, *8*(4), 351–362.
- Tallal, P., Merzenich, M., Miller, S., & Jenkins, W. (1998). Language learning impairment: integrating research and remediation. *Scandinavian Journal of Psychology*, *39*(3), 197–199.
- Thai, K. P., Mettler, E., & Kellman, P. J. (2011). Basic information processing effects from perceptual learning in complex, real-world domains. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 555–560). Boston, MA: Cognitive Science Society.
- Thelen, E. (2000). Grounded in the world: developmental origins of the embodied mind. *Infancy*, *1*, 3–30.
- Titchener, E. B. (1902). *A textbook of psychology*. New York: Macmillan.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208.
- Walk, R. D., & Gibson, E. J. (1961). A comparative and analytical study of visual depth perception. *Psychological Monographs: General and Applied*, *75*(15), 1–44.
- Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics*, *56*(5), 495–500.
- Warren, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. *Nature*, *336*(6195), 162–163.
- Wertheimer, M. (1959). *Productive thinking*. Westport, CT: Greenwood Press.
- Wertheimer, M. (1961). Psychomotor coordination or auditory and visual space at birth. *Science*, *134*, 1692.
- Wheeler, D. D. (1970). Processes in the visual recognition of words. *Dissertation Abstracts International*, *31*(2-B), 940.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*, 625–636.
- Wise, J. A., Kubose, T., Chang, N., Russell, A., & Kellman, P. J. (2000). Perceptual learning modules in mathematics and science instruction. In D. Lemke (Ed.), *Proceedings of the TechEd 2000 conference*. Amsterdam: IOS Press.
- Wurtz, R. P., & Lourens, T. (2000). Corner detection in color images through a multiscale combination of end-stopped cortical cells. *Image and Vision Computing*, *18*, 531–541.
- Zeki, S., Aglioti, S., McKeefry, D., & Berlucchi, G. (1999). The neurological basis of conscious color perception in a blind patient. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 14124–14129.
- Zhang, J. Y., Zhang, G. L., Xiao, L. Q., Klein, S. A., Levi, D. M., et al. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *Journal of Neuroscience*, *30*, 12323–12328.



Causation, Touch, and the Perception of Force

Phillip Wolff¹, Jason Shepard

Department of Psychology, Emory University, Atlanta, GA, USA

¹Corresponding author: E-mail: pwolff@emory.edu

Contents

1. Introduction	168
2. Force-based Accounts of Causation	169
2.1. Aristotle's Force-based Approach to Causation	169
2.2. Criticisms of Aristotle's Force-based Approach to Causation	171
2.3. Grounded Force-based Approaches to Causation	172
3. Grounding Forces in the Sense of Touch	174
3.1. Evidence for the Representation of Forces from Psychophysics	174
3.2. Evidence for the Representation of Forces from Weight Illusions	176
3.3. Evidence for the Representation of Forces from Neural Imaging	176
4. Evidence for a Force-based View of Causation from Work on the Perception of Causation	178
4.1. Michotte's Arguments against Force-based Accounts	179
4.2. How Michotte's Findings Indicate the Role of Forces in the Perception of Causation	181
5. The Sense of Force in Causal Perception and Induction	183
5.1. Forces in the Perception of Direct, Physical Causation	183
5.2. Forces in the Perception of Indirect and Social Causation	187
5.3. Correlations and the Sense of Force	189
6. The Feeling of Causation in the Absence of Mechanism: A Dual Process Approach	191
6.1. Explaining Causal Illusions in a Dual-process Framework	191
6.2. An Initial Test of the Dual Process View of Causal Understanding	193
7. Potential Problems for a Force-based View of Causation	196
8. Conclusions	198

Abstract

Much of the thinking on causation recognizes that it entails more than spatial-temporal contiguity or correlation, but it has been difficult to specify exactly what that extra component of thought is. In this paper, we argue that the representation of causal relations is based on the feeling of force as understood through the sense of touch. Grounding causation in people's sense of touch allows us to address the long-standing challenges that have been raised against force-based approaches to causation.

In support of our proposal, we review research on the perception of causation that provides support for a force-based view of causation. We also describe recent findings that establish a direct connection between people's impressions of causation and their sense of touch. We conclude by showing how a force-based view can be extended to handle the problem of how abstract causal relations are represented and acquired.



1. INTRODUCTION

Headlines in the popular press sometimes report correlations in a way that strongly suggests a causal relationship, such as “Eating Pizza Cuts Cancer Risk”¹ or “Eating fish prevents crime”². Such misrepresentations occur, no doubt, because causal claims indicate something deeper and more significant than correlational claims, but also because the exact nature of the difference between the two kinds of claims can be difficult to specify, thus affording a certain amount of wiggle room for artistic license.

In this article, we offer an account of how causation differs from correlation. In particular, we defend the thesis that the conditions that license causal attributions are not based on simple outward appearances, such as sequences of events occurring closely in space and time, but rather are based on the feeling of force as understood through the sense of touch. Thus, causal impressions are held to be grounded on more than the *kinematic* properties of an event—the entities, their motions, points of contact, and property changes—but rather on the perceived *dynamics* of an event, the forces and energies that bring about changes. As discussed below, the view that the sense of causation is based on force is arguably the first theory of causation put forth in recorded history, as well as the one that has been most heavily criticized. In offering a defense of forces, we will address some of the criticisms that have been raised, including the issues of the meaningfulness of forces and how forces might be acquired from experience. Our response to these criticisms will rest on the proposal that people understand forces as somatosensory impressions, i.e. in terms of their senses of touch, proprioception, and balance.

We begin by offering an account of how causation might be defined in terms of forces and then explain how this account fits within the history of proposals on causation. One of the themes that will emerge from this review is the issue of how causal relations might be acquired from

¹ BBC News, July 22, 2003, Retrieved from http://jfmuelller.faculty.noctrl.edu/100/correlation_or_causation.htm.

² The Observer, September, 14, 2003, Retrieved from http://jfmuelller.faculty.noctrl.edu/100/correlation_or_causation.htm.

experience. With respect to force-based theories, the issue centers on how people infer forces from the environment. As we will see, numerous studies have examined the encoding and storage of forces in the environment. After addressing the topic of acquisition, we focus on what prior research says about the representation of causation in terms of forces. Our search will include a close examination of the research on the perception of causation from collision events. With this background established, we describe recent findings that provide direct evidence for the proposal that casual relations are understood in terms of force, with force defined in terms of the sense of touch. We end with a discussion of how a force-based approach might account for the representation of abstract causal relations. In explaining the origins of causation, we arrive at an answer to the question of exactly how causation is more than mere correlation.



2. FORCE-BASED ACCOUNTS OF CAUSATION

In prior work, we have shown that causal relations can be understood in terms of configurations of forces (Wolff, 2007; Wolff & Song, 2003; Wolff, Barbey, & Hausknecht, 2010; Wolff & Zettergren, 2002). I refer to this account, which is based on Talmy's (1988) theory of force dynamics, as the dynamics model. According to the dynamics model, individual causal relations involve two main entities: an affector and a patient (the entity acted on by the affector). The theory holds that people specify causal relations in terms of configurations of forces acting on the patient. One of the forces acting on the patient is the force imparted on the patient by the affector. Another force is the force generated by the patient itself, or the patient's tendency to resist moving in a particular direction. The two forces can be added together to form a resultant force, which is then compared against an endstate vector in order to establish the patient's change in direction. The predictions of the theory have been tested in several studies (Wolff, 2007; Wolff & Song, 2003; Wolff et al., 2010). For current purposes, the main point is that this recent account of causation can be traced back almost 2500 years.

2.1. Aristotle's Force-based Approach to Causation

The idea that the concept of causation is based on force has its origins in ancient Greek philosophy. An initial first step was the atomistic influx model of Democritus (460–370 BC), which held that causation was produced by the transmission of an agent's substance—in the form of atoms—to the

patient (O'Neill, 1993). The atomistic influx model, in turn, appears to have influenced Aristotle's (384–322 BC) causal powers approach to causation. According to Aristotle, causation involves the transmission of a “form” from the agent to the patient (Marmodoro, 2007; Witt, 2008). For example, in a situation where a fire acts on a pot, the transmitted form would be heat. Aristotle emphasized that in a causal interaction, both the agent and patient have causal powers: the agent, the ability to transmit a form, and the patient, the capacity to receive the change. It is interesting to note that throughout Greek scientific literature, the word for power (and sometimes force) was *dynamis* (Jammer, 1957), which eventually gave rise to the modern day word *dynamics*, the branch of mechanics that deals with forces and their relation to motion. In Greek, *dynamis* meant “strength” or “power”, but also “ability” or “faculty” (Witt, 2008). Aristotle differentiated two types of causal powers: active and inactive. Inactive powers were understood as tendencies or potentialities for some kind of actuality³. As described below, this aspect of Aristotle's theory of causal power was carried over into his thinking about forces.

In addition to a causal powers theory of causation, Aristotle was one of the first philosophers to make an explicit link between causation and force (Jammer, 1957). It has been stated that Aristotle viewed forces as a particular type of cause (Heidegger, 1995), namely, the type associated with physical pushes and pulls (Jammer, 1957). Evidence for this view can be found in many places, including multiple references to forces being causes in Aristotle's eighth book of his *Physics* (1999):

Of intrinsic motions, some are caused by the thing itself, some by another thing, and some happen by nature, and others happen by force and contrary to nature.
(p. 9)

In “The Eudemian Ethics”, we see that Aristotle also saw forces as causally relevant to the actions of living things and that he viewed them much like causal powers since they interacted with tendencies (2011):

Similarly, with living things, including animals, we see them being acted on by force, and also acting under force, when their motion is caused by an external agent against their intrinsic tendency. (p. 27)

Indeed, Aristotle (2011) plainly states that forces can serve as psychological causes in a manner analogous to physical forces.

³Talmy's (1988) theory of force dynamics shares a number of features with Aristotle's theory of causal powers, in particular, the notion that both the agent and patient have intrinsic tendencies and play a role in a causal interaction.

It has already been said that these people seem, exceptionally, to act both voluntarily and by force, the reason being a certain similarity to the kind of force that we speak of also in connection with inanimate things. (p. 28)

2.2. Criticisms of Aristotle's Force-based Approach to Causation

Aristotle's ideas about causation dominated thinking on the topic for the next 2000 years. However, during the fifteenth and sixteenth centuries, it came under wide criticism. A quick review of these criticisms is worthwhile because some of the same arguments are still put forth today (e.g. Cheng, 1997; Cheng & Novick, 1991, 1992; Schulz, Kushnir, & Gopnik, 2007; Woodward, 2007).

For philosophers in the fifteenth and sixteenth centuries, Aristotle's account of causation fell short because the notion of force was too mysterious to be useful. To understand this perspective, it needs to be recognized that during this time, many philosophers were attracted to the idea of natural mechanism, that is, the view that living and other natural things behaved the way they did because they were like machines (Ott, 2009). As machines, their behavior could be explained in terms of their mechanical properties, that is, in terms of local interactions of parts, just as the behavior of a clock could be explained in terms of the size, shape, position, and movement of its gears. The Aristotelian view of causal powers and forces conflicted with the mechanistic view because it was unclear how these notions could be defined in mechanistic terms. For example, it was unclear how defining gravity as a force improved our understanding of gravity, unless the notion of force could be grounded in terms of local interactions of parts. Explanations based on causal powers or forces were viewed as little more than appeals to the occult.

Some modern philosophers, such as John Locke (1632–1704) and Robert Boyle (1627–1691), tried to combine Aristotelian ideas with mechanistic thinking (Ott, 2009). Others, including René Descartes (1596–1650), Nicolas Melebranche (1638–1715), and David Hume (1711–1776), ruled out the possibility of causation in terms of powers or forces altogether. For Hume (1748/1975) in particular, the problem with force was that it could not be linked to any internal or external sensory impressions. Hume (1748/1975) acknowledged that after many repetitions of conjunctions of objects or events, people would develop an expectation that could be interpreted as a power or force, but this power or force could not serve as the basis of causation because it resulted from thoughts projected onto experience rather than from experience projected onto thought.

Newton's (1643–1727) theory of force gave modern philosophers pause (Jammer, 1957). On his account, the notions of causation and force were connected. In his *General Scholium*, Newton wrote that causes were forces that changed the course of events, specifically, “The causes by which true and relative motions are distinguished, one from the other, are the forces impressed upon bodies to generate motion” (Newton, 1687, I: 14). Newton inspired other force-based theories, including those developed by Martin Knutzen (1713–1751) and Immanuel Kant (1724–1804). What Newton's theory suggested was that forces might be instantiated independently of the mind, which must hold if they are to serve as the basis of causation. At issue, then, is whether they could be perceived directly. According to Hume (1748/1975), they could not because they left no impression on the senses.

2.3. Grounded Force-based Approaches to Causation

Several proposals have challenged Hume's claim that forces cannot be sensed, the common thesis being that forces can be sensed if they are linked to our will or haptic sense. This idea was first proposed by Thomas Reid (1710–1795), who suggested that the ideas of force, power, and causation were derived from our conscious awareness of voluntary actions. According to Reid, “It is very probable that the very conception or idea of active power and of efficient causes is derived from our voluntary exertions in producing effects...” (Reid, 1788/2010, p. 250; see also Jammer, 1957). A similar view was offered by Maine de Biran (1766–1824), who, like Reid, argued that the prototype for our idea of force is found in our own will (Jammer, 1957; Piaget, 1930/1969; Truman, 1904). For de Biran, the concept of cause is based on the inner consciousness of force that stimulates a voluntary bodily motion or mental process as well as the kinesthetic sensations that accompany muscular contraction. Piaget (1930/1969) held the related belief that forces consisted of schemas built up from muscle experiences and the accompanying sense of effort. Piaget viewed his theory as partially consistent with de Biran's in that they both viewed force as having an internal origin that could later be attributed to things in the external world.

In the current literature, an account related to all of these hypotheses has been proposed by White (1999, 2006, 2009, 2012a, 2012b). According to White (1999, 2006), the idea of causation originates from actions on objects. Such actions provide two kinds of input. The first is the experience of motor activity; the other is the haptic sensations produced by pressure sensors in the skin and sensation from bodily position, weight, muscle tension and movement (i.e. kinaesthesia). White emphasizes that both

motor activity and haptic sensation are required to establish knowledge of causation. White further argues that actions on objects lead to the formation of schemas specifying forces. These schemas are important because when people see patterns of motion, these patterns are matched to the schemas of force stored earlier. Thus, it is people's direct experiences acting on objects that allow them to infer forces in events perceived visually (White, 2012a; see also Piaget, 1930/1969). In more recent work, White (2011, 2012a) concludes that the perception of causation is not directly based on the perception of forces, though he still holds that the perception of forces is based on people's experiences acting on objects and that collision events give rise to the perception of force.

In all of the proposals discussed so far concerning how forces may be grounded in the body, the idea has been that when we perceive forces in the world, we adopt the role of agent over the role of patient. In other words, we apply our own personal experiences to entities in the external world from the point of view of the active entity rather than the inactive one (White, 2006, 2012a). A recent proposal by Fales (1990) adopts the alternative possibility. As in the other proposals, Fales (1990) argues that causation is based on force and that our notions of force have a sensory basis. Fales (1990), however, emphasizes that our notions of force are based primarily on experiences in which our bodies are acted on by other entities. In particular, he highlights the role of the tactile perceptions that accompany impingement upon our bodies as well as our kinesthetic sense and our sense of balance (1990). Our kinesthetic sense records sensations associated with the relative motion between different parts of our bodies produced by the extension of skeletal muscles. Our sense of balance depends on the vestibular apparatus in the inner ear and is designed to sense generalized forces such as gravity and centrifugal forces. Fales (1990) points out that our tactile and kinesthetic senses may be sensitive to different forces, or to the same force. Importantly, the sense of force experienced through these different senses can be integrated using vector algebra to compute resultant forces specifying magnitude, direction, and point of application. As we will discuss below, recent experiments support Fales' (1990) patient-oriented theory over White's (1999, 2006, 2012a), Reid's (1788/2010), Maine de Biran's (Jammer, 1957; Piaget, 1930/1969; Truman, 1904) and Piaget's (1930/1969) agent-oriented theories.

Recent years have witnessed a resurgence of interest in the idea that causation might be based on causal powers (Ahn & Bailenson, 1996; Ahn, Kalish, Medin, & Gelman, 1995; Aronson, 1971; Bullock, Gelman, & Baillargeon,

1982; Dowe, 2001; Fair, 1979; Harre & Madden, 1975; Salmon, 1994, 1998; Shultz, 1982). This renewed interest has resulted in demonstrations of how force-based representations allow us to differentiate causation from noncausation and various other types of causal categories (Ahn & Kalish, 2000; Bigelow et al., Bigelow & Pargetter, 1990; Leslie, 1994; Strawson, 1987; Talmy, 1988; Wolff, 2007; Wolff & Zettergren, 2002). Nevertheless, the notion that forces play a role in causal understanding continues to attract significant opposition. Arguably, the strongest criticism against force-based accounts is that originally raised by Hume (1748/1975), namely, that the notion of force cannot provide a basis for causation because it cannot be sensed. Recent research indicates that Hume overstated his case. While it is true that forces are invisible, vision is not our only sensory modality. Once we consider the potential contributions of the other senses, most notably touch, it becomes clear that people's sensory experiences are not as deficient as Hume claimed.



3. GROUNDING FORCES IN THE SENSE OF TOUCH

The brain, in fact, devotes a fairly large amount of real estate to the processing of forces. The system responsible for determining whether forces have been applied against the body is the somatosensory system. Anatomically, the system is divided into two parts. The postcentral gyrus (or anterior parietal cortex), which consists of Brodmann areas 3, 1, and 2, is the location of the primary somatosensory cortex (SI) and the parietal operculum (the upper bank of the lateral sulcus) is the location of the secondary somatosensory cortex (SII) (Keysers, Kaas, & Gazzola, 2010). Both cortices process touch and proprioception. Touch is the sense by which pressure exerted on the skin is perceived, pressure being a function of force, namely, force divided by surface area. Proprioception (or kinesthesia) is the sense through which we perceive the position and movement of our body, including our sense of equilibrium and balance, senses that depend on the notion of force (Jones, 2000).

3.1. Evidence for the Representation of Forces from Psychophysics

There is much evidence that the somatosensory system, in particular the tactile system, is fairly adept at distinguishing force magnitudes. This finding emerged, for example, in a study by Wheat, Salo, and Goodwin (2004), in which participants estimated the magnitudes of forces applied to their index fingers. The researchers observed a nearly linear relationship between participants' estimates and the actual magnitude of the force acting on

their fingers (see also Jones & Piateski, 2006). Related research has shown that people are able to distinguish force directions. For example, Panarese and Edin (2011) asked participants to discriminate the directions of forces applied to the index finger and found that they were able to discriminate forces that differed by only 7.1°.

Beyond basic discrimination, abundant evidence indicates that people store information about forces. Indeed, the storage of forces is revealed in common everyday tasks. Many of us, for example, have had the experience of reaching for a suitcase or box and over-lifting it because we thought it was full when, in fact, it was empty (Reinkensmeyer, Emken, & Crammer, 2004). Such events suggest that we estimated the weight of the suitcase, a type of force, before we lifted it, and because we estimated wrongly, we generated greater-than-necessary forces. Experimental evidence for the storage of forces comes from studies that put people in “force fields” and then observe their subsequent motor actions. For example, in a classic study by Shadmehr and Mussa-Ivaldi (1994), participants were instructed to move their hand from one point to another while holding onto a robotic arm (i.e. a manipulandum). The robotic arm was programmed to generate forces that pushed the person’s hand away from the target location. With repeated practice, people learned how to overcome the pressure of the robotic arm and to reach straight for the intended target. The key finding was the appearance of an aftereffect once the force field (robotic arm) was removed: people’s arm trajectories were distorted in the opposite direction of the previously applied force. The result implies that people had internalized the force field. Similar findings have been observed in conditions of microgravity, that is, when people are placed into the weightless condition of parabolic flight and asked to reach for certain targets (Papaxanthis, Pozzo, & McIntyre, 2005). Changes in the trajectories of their arms imply that people factor into their motor plans the forces of gravity and inertia. Another type of experiment examining lifting behavior has revealed evidence for the role of forces in people’s motor actions. For example, it has been found that people generate larger gripping and lifting forces when they pick up larger objects than smaller objects (Gordon, Forssberg, Johansson, & Westling, 1991). Additional research on lifting shows that people can store information about more than one force and then combine these forces. In a fascinating study by Davidson and Wolpert (2004), people learned the forces needed to lift two objects independently. The two objects were then stacked together. Pressure sensors on the objects showed that the amount of pressure they applied to the stack of objects was a weighted sum of the pressures they had applied to the objects when they were lifted independently.

Evidence for the representation of forces in these motor planning studies has been attributed to the formation of an internal model that represents the dynamic properties of the environment (see also [Conditt, Gandolfo, & Mussa-Ivaldi, 1997](#); [Hinder & Milner, 2003](#); [Imamizu, Uno, & Kawato, 1995](#); [Kawato, 1999](#); [Milner & Franklin, 2005](#); [Ohta & Laboissière, 2006](#)).

3.2. Evidence for the Representation of Forces from Weight Illusions

Research examining people's judgments of weight have found that inferences about forces are based on several kinds of cues. For example, [Hamilton, Joyce, Flanagan, Frith, and Wopert \(2007\)](#) observed that people used the fine details of a lifter's kinematics in their estimates of weight. In their study, participants watched videos of a person picking up boxes of different weight and then estimated the weight of a box on a 1 to 100 scale. Participants' weight estimates were found to be a function of the duration of the lift phase, transportation phase, and grasp phase (see also [Shim & Carlton, 1997](#)). Other cues people used to judge an object's weight were its size and density. They tended to assume that large and dense objects weighed more than small and airy objects ([Walker, Francis, & Walker, 2010](#)). Research on various types of weight illusions tell us that these cues are used spontaneously in people's estimations of weight. For example, in the size-weight illusion, when people are asked to estimate the relative weight of a large object and a small object that in fact weigh the same, they will mistakenly perceive the smaller object as weighing more than the larger object ([Flanagan & Beltzner, 2000](#); [Kawai, Henigman, MacKenzi, Kuang, & Faust, 2007](#)). The phenomenon is typically explained as resulting from a mismatch between expectations and actual sensory experience. People expect the small object will weigh less than the large object, but when they learn this isn't the case, they mistakenly over-estimate the weight of the smaller object. In a related illusion, the weight of objects made of low-density materials (e.g. Styrofoam) are perceived to weigh more than those made dense materials (steel) even though their actual weights are exactly the same ([Ellis & Lederman, 1999](#)). The key point for our purposes is that the perceptual-conceptual system spontaneously estimates forces on the basis of cues such as size and material.

3.3. Evidence for the Representation of Forces from Neural Imaging

In all of the research on people's ability to represent force discussed so far, participants were asked to either perform a motor action or provide

a judgment. One question left open by this research is whether people make spontaneous inferences about forces even when an explicit action or judgment is not required. Recent work using neural imaging suggests that they do. The surprising finding from this research is that the somatosensory cortices become active not only when a person is touched but also when they observe touching. In a study by [Keysers et al. \(2004\)](#), participants were either touched on their legs or viewed movies of other people or objects being touched. They found that the SII was activated in all conditions. In particular, activity in SII was observed both when the person was directly touched on the leg as well as when they saw another person being touched on the leg. Moreover, when the legs in the video were replaced by paper towels, activity in SII was still observed when there was touching, suggesting that the activity in SII corresponds to a relatively abstract notion of touching. Interestingly, in the observation conditions, [Keysers et al.](#) did not find activity in SI. As noted by [Keysers et al.](#), the finding of activity in SII is consistent with recent interpretations of SII as a site of integration between somatosensory information and information from other senses, like vision.

[Keysers et al.'s \(2004\)](#) findings have been replicated and extended in several other studies. For example, [Blakemore, Bristow, Bird, Frith, and Ward \(2005\)](#) found that observing another person touch his or her face resulted in activity in the face region of the viewer's SI. [Ebisch et al. \(2008\)](#) presented participants with videos showing intentional and accidental touching occurring between animate and inanimate objects. The videos involved scenes in which a person touched another person or a chair, or a branch touched a person or a chair. As with [Keysers et al. \(2004\)](#), [Ebisch et al. \(2008\)](#) found activity in SII for all conditions. Interestingly, they found some activity in SI/BA 2 for the videos depicting intentional touching (i.e. the videos in which the agent was a person). In a study using magnetoencephalography, [Pihko, Nangini, Jousmäki, and Hari \(2010\)](#) obtained a similar pattern of findings. The participant's hand was touched by the experimenter or the participant observed the experimenter touch her own hand. As in previous studies, the somatosensory cortex was activated in both the directly experienced and observed conditions. However, unlike in previous studies, [Pihko et al. \(2010\)](#) observed activity in the SI only. Also, interestingly, [Pihko](#) observed that the activity in the observed condition was 7.5% of the activation in the directly experienced condition. Finally, [Meyer, Kaplan, Essex, Damasio, and Damasio \(2011\)](#) pursued a very different approach to the question of whether touch activates the somatosensory cortex by probing the informational content of the activity in that region. Participants

watched people handle various everyday objects. As in previous studies, it was found that watching touching resulted in activity in the somatosensory cortex, specifically SI. Especially impressive, using multivariate pattern analysis, Meyer et al. (2011) were able to predict which of the several objects was being handled based exclusively on the pattern of activity in SI.

One question these studies did not address concerns how the link between vision and touch is formed. One possibility is that the association is built into the architecture from birth. Alternatively, as proposed by Keyers et al. (2004), it may result from ordinary associative learning. When people see themselves being touched, this visual sensation will overlap with the somatosensory sensation of being touched. With repetition, this association will be built up so that people experience a somatosensory sensation of being touched in response to a visual stimulus of touching, even in the absence of actual physical touch.

Under the assumption that forces are processed, at least in part, in the somatosensory cortex, the studies reviewed above offer strong evidence that people are able to represent forces and that they spontaneously engage in the encoding of forces from visual information. Thus, in contrast to the claims of Hume, people's sensory experiences do appear to include forces.



4. EVIDENCE FOR A FORCE-BASED VIEW OF CAUSATION FROM WORK ON THE PERCEPTION OF CAUSATION

The previous section established that people automatically infer and store forces from the environment, but such abilities do not necessarily imply that their notion of causation is based on forces, just that it *could* be based on forces. In the following sections, we examine the evidence in support of a force-based view of causation. We begin by examining the extent to which past research offers evidence in support of the view that people's representations of causal relations are based on forces. In particular, to what extent does the launching event give rise to the perception of forces, which in turn leads to a causal impression?

Michotte (1946/1963) provides a mostly contradictory answer to this question. The bulk of his work emphasizes the importance of kinematics over dynamics. This emphasis comes through especially strongly in experiments looking at "paradoxical cases", situations in which the essential conditions for the causal impression are fulfilled but in such a way that the stimulus properties of the event conflict with everyday experience and the

laws of mechanics. Michotte cited such cases to argue that the stimulus properties that give rise to the causal impression must be innate because they could not be learned from experience. In fact, a close examination of such cases (see below) shows that they are not so much at odds with the laws of mechanics as Michotte claimed. Moreover, if we consider the more typical cases giving rise to the causal impression, we see that they appear to provide strong support for the claim that the causal impression is based on forces because the stimulus characteristics of these events are exactly those of events in which forces are produced. Indeed, in the conclusion of his book, Michotte seems to change stories rapidly when he asserts “the causal impression is *the perception of the work of a mechanical force*” (p. 228). Interestingly, there is a way in which the two parts of Michotte’s story can be unified. As noted in the previous section, certain stimulus properties, such as size and material, can serve as cues to force. Such cues are not infallible, just relatively reliable indicators of force. Similarly, the stimulus conditions identified by Michotte may serve as cues to force rather than as direct triggers of the concept of cause. A review of some of Michotte’s paradoxical cases will offer support for such an account.

4.1. Michotte’s Arguments against Force-based Accounts

One reason why Michotte felt that the causal impression was based on a particular stimulus pattern and not the laws of mechanics was that people sometimes reported perceiving causation in events that he viewed as physically impossible. For example, people reported perceiving causation in situations in which objects A and B were moving in the same direction, A faster than object B, and after A hits B, A stopped and B slowed down (Michotte, 1946/1963, p. 71). On the basis of such results, Michotte concluded that the causal impression was not based on past experience with the world, which of course, honors dynamics: had people referred to past experience, they would have expected object B to speed up, not slow down, after being hit. However, while such the experimental sequence of events may be unusual, it is not necessarily at odds with a force-based account of causation. Friction can change dramatically over the course of an object’s movement, as when a ball rolls off an asphalt road and onto a gravel driveway. Michotte’s “impossible” event is not, in fact, impossible in the world, and so his finding does not necessarily rule out the role of forces in the perception of causation.

Another of Michotte’s arguments for the independence of the launching effect and the laws of mechanics is that sometimes the causal impression failed to obtain for trajectories that people experience in the real world. In

support of this point, Michotte conducted several Experiments (34 and 35) in which object A hits object B directly and B traveled at an angle away from its expected straight-line path. The degree of deviation from B's expected straight path ranged from 25 to 90°; as the size of the angle increased, the causal impression grew weaker. Recent studies (Straube & Chatterjee, 2010; White, 2012b) have replicated this finding. Michotte points out that this result is at variance with our real-world experience in which two colliding objects can travel at angles (beside 180°) and still be viewed as causal (e.g. billiards, marbles). However, Michotte's collision events were quite different from those involving billiard balls and marbles. As noted by White (2012b), in the real world, B's direction of movement depends not just on the direction of A, but also on where B is hit, that is, on its point of contact. If an object is hit from below its center of mass, a force view predicts that the object will move upwards, in a direction that differs from A's, not straight ahead. In Michotte's experiments, A hits B head on, and so a force-view would predict that B's direction should be straight ahead, but what people saw was B moving away at an angle. Given that such direction is at variance with the forces involved in the situation, it is not surprising that people's causal impressions decreased as the angle of departure increased. Indeed, in an experiment reported in White (2012b), people's causal impressions were high for events involving angles, so long as the direction conformed to the direction that would be expected from A's direction and A's and B's point of contact. As White (2012b) notes, this result contradicts Michotte's hypothesis that the causal impression depends on continuity of motion between A and B, and instead supports the view that the causal impression depends on people's real world experiences with collision events.

A third argument raised by Michotte that the causal impression was not tied to the laws of mechanics is that people experience the causal impression even when the objects involved are spots of light, shadows, or lines painted on a rotated disk. In other words, people perceive causation while also knowing that such causation does not occur in the real world (1946/1963, pp. 84–85). However, a force-based approach to causation does not imply that people cannot be subject to illusions of causation. A particular configuration of forces will produce only one kinematic pattern, but a single kinematic pattern is potentially consistent with more than one configuration of forces. This asymmetry explains why causal illusions can sometimes occur: people may infer the wrong configuration of forces from a particular kinematic pattern. This is especially likely when the actual forces driving the kinematics are obscured, as in the case of Michotte's launching

events. Further, the process of inducing forces is likely to be at least partially automatic (Runeson & Frykholm, 1983), so causal illusions may occur even when the inferred configuration of forces is inconsistent with prior knowledge of the situation.

It should also be noted that while Michotte claimed that the launching event was largely independent of the shape and size of the objects involved in the event, more recent research indicates that object properties do, in fact, influence the perception of causation (for a review see Saxe & Carey, 2006). A convincing example of the importance of object properties on the impression of causation was demonstrated in a study by Kotovsky and Bailargeon (1998), in which 5.5- and 6.5-month-old infants were shown an event in which a cylinder rolled down a hill and hit a “bug”, after which the bug moved to the center of the stage. Once the infants were habituated to this event, they were shown the same event again, except that the cylinder was replaced with either a smaller or a larger cylinder, and the bug moved further across the stage. One of the key findings was that the 6.5-month-olds looked longer at the trials involving the small cylinder than at those with the large cylinder, suggesting that they were surprised to see a larger effect follow from a smaller causal object. This result makes sense if infants’ causal impressions are based, at least in part, on forces: smaller objects cause smaller forces and hence, smaller effects; any other pattern is viewed as surprising.

4.2. How Michotte’s Findings Indicate the Role of Forces in the Perception of Causation

Michotte emphasized that the causal impression was not a mere copy or reproduction of what goes on in the real world, but the main findings of his research program indicate just the opposite. For example, Michotte observed that the causal impression disappeared when there was a temporal delay of around 150 ms between the moment objects A and B made contact and the moment B began to move. This finding is easily explained by a force-based account of causation. When object A hits object B, the force imparted on B is instantaneous. If object B begins moving well after it is hit, its movement cannot be due to the force imparted by object A. The importance of temporal contiguity in the perception of cause has been replicated in a number of studies (Morris & Peng, 1994; Oakes & Kannass, 1999; Schlottmann & Anderson, 1993; Schlottmann & Shanks, 1992; White, 2010). Another finding of Michotte’s is that the perception of causation is strongest when object A makes physical contact with object B. This finding is also consistent with a force-based approach, since contact forces cannot

exist unless objects make contact with one another. The effect of physical contact on the causal impression has also been demonstrated in a several studies (Kotovsky & Baillargeon, 2000; Schlottmann, Ray, Mitchell, & Demetriou, 2006; Spelke, Phillips, & Woodward, 1995).

Another phenomenon associated with the causal impression is the *radii of action*. The radii of action are the portions of the paths traveled by A and B that appear to be relevant to the impression of causation. When B travels beyond A's radius of action, it appears to be moving on its own, not as a consequence of the collision. Michotte found that object B's radius of action increased with the speed of object A. Michotte was unable to offer an explanation for the phenomenon because whether B remained within the radius of action or traveled beyond it had no consequence for event's continuity of motion, the hypothesized source of the causal impression. In contrast, force-based approaches to causation offer a natural explanation: as object A's speed increases, the force it imparts on B increases, and, in turn, so does the distance B travels as a consequence of A's impact (for a related proposal, see Hubbard & Rupel, 2002).

Finally, as noted above, according to Michotte, the causal impression should be strongest when the two parts of a launching event constitute a single continuous movement, whereby the motion of the first object extends into the second and creates an "ampliation of motion." According to this hypothesis, any differences in velocity between the first and second objects should decrease the causal impression, because any difference in velocity makes the sequence of events less continuous. However, in contrast to this prediction, Michotte found that the causal impression was stronger when the speed of object B was slower than that of object A. Specifically, in Experiments 15 and 39, people reported a much stronger causal impression when the ratio in speed of objects A and B was 4:1 than when the ratio was 1:1. This result is consistent with a force-based approach to causation. The reason why the second object moves less rapidly than the first is because at the point of contact there is loss of energy. Moreover, under the assumption that B's movement is due to external forces, B should ultimately slow down as it is acted on by friction with the surface. When object B's speed is the same as object A's, force-based accounts predict that the causal impression should be weaker because it suggests that some other forces must be involved in the production of B's movement.

In sum, research on the launching event supports the thesis that the notion of causation is ultimately based not on outward appearances, but rather on the notion of force. Once we conceptualize causation in terms

of force, we are able to explain why the perception of causation depends on spatial and temporal contiguity in the launching event. We are also able to explain why the perception of causation is affected by differences in the speed of the two objects. In sum, a force-based approach allows us to make better sense of the phenomena surrounding the launch event. That said, the current literature does not offer the kind of evidence needed to make the claim that causal relations are associated with the representation of forces. The research described in the next section takes an initial step toward filling this gap.



5. THE SENSE OF FORCE IN CAUSAL PERCEPTION AND INDUCTION

The proposal that causation is based on force has implications for the perception of causation. The proposal implies that when people see causal events, they should simultaneously infer forces. In the experiments described below, we tested this possibility (Wolff, Ritter, & Holmes, in preparation) by examining whether “seeing a force” had an effect on “feeling a force.” In order to test this prediction, we needed to be able to impart precisely timed forces and measure people’s reaction times to these forces. This was accomplished using a haptic controller device. A haptic control is essentially a small robotic arm. Like a mouse, you can push it around, but unlike a mouse, it can push back. The arm has impressive capabilities. It can be used to “feel” virtual surfaces that are bumpy, sticky, smooth or rough. For the purposes of the following experiments, we needed the arm to do one simple thing: impart a force at an exact point in time against people’s hands. The controller had a small button that people could press to indicate that they experienced a force. The controller was programmed using widely available C++ libraries.

5.1. Forces in the Perception of Direct, Physical Causation

In the first three experiments, we focused on events involving physical forces. As described in greater detail below, each experiment included three conditions, which were run between participants with 25 participants in each condition, for a total of 75 participants in each experiment. Frames from the animations used in these experiments are shown in the first three pictures of Figure 5.1. In each experiment, participants saw both causal and noncausal animations. The two kinds of animations were designed to be as similar as possible. In Experiments 1–3, the causal animations depicted

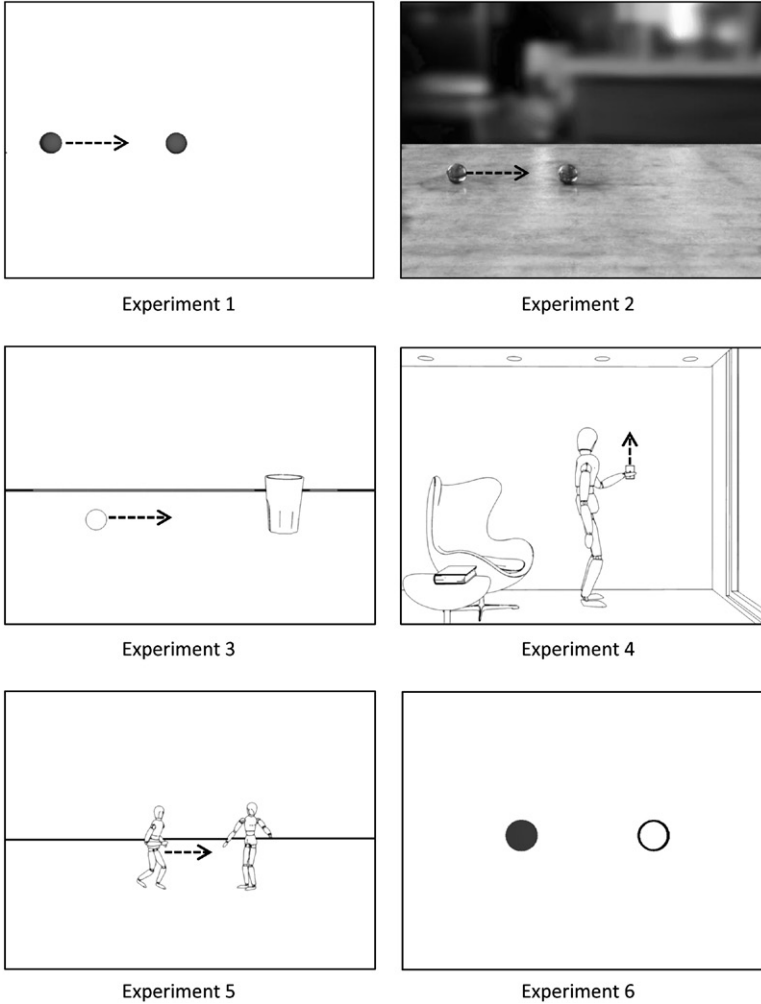


Figure 5.1 Frames from the causal animations used in experiments 1 through 6. In Experiment 1, shown in the top left panel, the ball on the left hits the ball on the right, sending it into motion. In Experiment 2, shown in the top right panel, the motions were the same as in Experiment 1, except the surfaces were near-photorealistic. In Experiment 3, the surfaces were the same as in Experiment 2, except the second marble was replaced with a glass cup that shattered. In Experiment 4, the person on the left approaches the one on the right, who directs that person to turn to the right; the animation depicted a near-photorealistic scene. In Experiment 5, the person on the left approaches the one on the right, who directs that person to turn to the right; the animation depicted a desert scene. In Experiment 6, there is no motion, rather, the circle on the left turns solid, and then a few moments later, the circle on the right turns solid.

collision events. In Experiment 1, in particular, the background was black and a red ball hits another red ball, sending it into motion. The noncausal variant of this animation showed a single ball move across the screen at the same exact rate as the balls in the causal animation. In Experiment 2, causal and noncausal animations were exactly the same as those used in Experiment 1, except that the animations were rendered using near-photorealistic surfaces. The balls were blue marbles that rolled on top of a marble countertop, and an out-of-focus background suggested various kinds of kitchen appliances, a window, and a sink. The causal and noncausal animations in Experiment 3 were much the same as those used in Experiment 2, except that the second marble in the causal animation was replaced with a small, clear glass cup that shattered upon impact. In the noncausal version of this change-of-state event, the glass was removed and the ball traveled the exact same path traveled in the causal version of the event.

In all three experiments, the trial structure was the same. In each trial, participants held the haptic controller and saw the same animation four times. The first three times, the animation played at different speeds, randomly chosen, such that the animation lasted 540, 1440, 2340 or 3240 ms. We showed the animation several times in order to “build up” the sense of force. The fourth time the animation played, it lasted an intermediate amount of time, 1800 ms. At the end of the last animation, the last frame of the animation was paused, and the haptic controller moved 100, 200, 300, 400, or 500 ms after the onset of the last frame. When they felt the controller move, participants were to press a button on the controller to indicate that they had felt a force. The movement times were varied so that participants could not predict exactly when the controller would move. The force generated by the controller was very small, specifically 1.5 N for 20 ms, which, phenomenologically, produced a very faint impression on the hand, but clearly above the sensory threshold for touch. There were 20 practice trials, half causal, half noncausal, and 40 experimental trials, again, half and half.

The main prediction was that “seeing” a causal event would affect people’s speed at detecting a force, most likely in the direction of decreasing their reaction times, giving rise to a type of “priming” effect. While such a result would support the hypothesis that people infer forces when they see causal events, it could also be consistent with several other less interesting possibilities. In particular, such a result could arise if the causal events were better predictors of the onset of the force than the noncausal events, despite our efforts to discourage such predictive processes. Alternatively, the causal

events could have been more interesting than the noncausal events, hence increasing people's arousal level and ultimately decreasing their response times.

To guard against such possibilities, two control conditions were added to the experiments. These conditions were exactly the same as that described above except that, instead of a force, participants were subject to either an auditory or visual signal. In the auditory signal control condition, participants heard a brief electronic sound through a set of earphones. In the visual signal control condition, a dot was briefly flashed above the last object at the very end of the animations. In both conditions, the participants' task was to press the button on the haptic controller as soon as they either saw the dot or heard the sound. The controller did not move in either of these conditions. These conditions are important because if effects are found in the force condition, but they are due to uninteresting reasons, such as predictability or arousal levels, then the same effects should also be observed in the visual and auditory conditions. In contrast, if there is an effect in the force condition and not in the auditory and visual control conditions, the overall pattern of results will suggest that the effect of seeing a causal event is specific to the touch modality, just as we are predicting. In sum, our main prediction was that people would be faster to detect a force after seeing a causal than noncausal event, and that this effect would only be observed in the force condition.

The predicted results were obtained in all the experiments. The results from Experiment 2 are representative of all of the experiments. As shown in Figure 5.2, people responded faster to a force against their hand after seeing

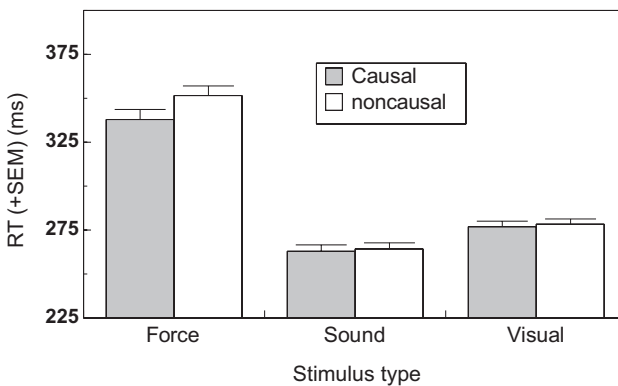


Figure 5.2 Results from Experiment 2 showing reaction time to respond to a force, sound, or visual stimulus after watching a causal or noncausal animation. Error bars indicate standard errors of the mean.

a causal than noncausal event, $t(24) = 2.38$, $p = 0.025$. Importantly, this difference cannot be explained as an effect due to greater predictability of the signal after seeing the causal than noncausal event, or to greater arousal after seeing the causal versus noncausal event, because times to respond to a sound or visual stimulus did not differ after watching a causal or noncausal event. Rather, watching causal events only had an effect on speed to detect a force.

In Experiment 2, the reaction time to respond to a force after watching a causal animation minus the time to respond to a force after watching a noncausal animation was -13.5 ms; in other words, participants were 13.5 ms faster to respond to a force after viewing a causal animation than a noncausal animation. There was no evidence for such a difference in the sound ($D = -1.35$) or visual ($D = -1.47$) conditions. The same set of differences for the other experiments are listed in Table 5.1, and as can be seen, the pattern was the same across the other experiments. In Experiments 1 and 3 as well, people were statistically faster to report detecting a force after watching a causal animation than after watching a noncausal animation, but there was no evidence for a similar difference when detecting a sound or visual signal. Altogether, the results from Experiments 1–3 suggest that when people see a causal event involving physical forces, they perceive these forces in a rather direct manner.

5.2. Forces in the Perception of Indirect and Social Causation

It could be argued that the events used in Experiments 1–3 were special in that people had direct access to the underlying mechanism of the causation, which involved direct physical contact. In many everyday causal relations, we do not have access to the underlying mechanisms behind the causation. A reasonable question, then, is whether forces are experienced in the case of events in which the underlying physical mechanism is hidden. This question was examined in Experiment 4. The procedure, trial structure and experimental design were the same as in the previous experiments. The only difference was that the experiment used materials in which the causal mechanism was hidden; specifically, in the causal condition a person flipped a switch and the ceiling lights came on and in the noncausal condition a person flipped a switch and the ceiling lights remained on. Hence, the two animations ended up exactly the same, namely, with the lights on. The animations used in this experiment were near-photorealistic, so people could easily differentiate when the lights were on or off. As shown in Table 5.1, the pattern of results was the same as in the previous experiments. Participants were faster

Table 5.1 Difference in RT to Indicate Detecting a Force, Sound or Visual Signal after Watching a Causal versus Noncausal Animation for Experiments 1–5 in Milliseconds

	Stimulus type		
	Force	Sound	Visual
Experiment 1: schematic collision	-9.2	0.21	-0.72
Experiment 2: realistic collision	-13.5	-1.35	-1.47
Experiment 3: shattering	-7.6	-1.01	-2.019
Experiment 4: turning on a light	-14.8	1.28	-3.58
Experiment 5: social causation	-15.1	-0.070	-2.49

to respond to a force after watching the causal than noncausal event, but there was no evidence for a similar difference in the sound and visual conditions. The results from this experiment suggest that people experience forces in causal situations even when the physical forces are hidden.

In Experiment 5, we examined an even more abstract type of causation, that of social causation. The procedure, trial structure, and design were the same as in the previous experiments. In the cause condition, participants saw a person direct another person to change their path of motion. In the non-causal condition, participants saw a person travel the same path of motion as in the causal condition, but without another person directing them to change their path. As in previous experiments, the animations included photorealistic surfaces, in this case, a desert scene with a large open sky. As shown in Table 5.1, the pattern of results mirrored that of the earlier studies. Participants were significantly faster to report a force after watching the causal events than the noncausal events, and this difference only occurred in the force condition. The results imply that social causal events are experienced in a manner similar to how people experience physical causal events.

Across a relatively wide range of situations, we found that causal events were associated with the experience of force. Interestingly, the results suggest that the perception of force affected people's touch sensitivity and not their motor planning. Had the perception of force affected motor activity we should have seen faster RTs after seeing causal than noncausal events not only in the force condition, but also in the sound and visual conditions. The effect should have been present in all three conditions because in all three conditions participants had to engage in the motor activity of pressing a button. The results suggest, then, that when people perceive forces from visual stimuli, they "empathize" with the object that suffers the effect, the patient, which sensitizes their sense of touch, and not with the object that brings about the event, the agent. Thus, our results are more consistent with

Fales' (1990) patient-oriented account of force perception than White's (1999, 2006, 2012a), Reid's (1788/2010), Maine de Biran's (Jammer, 1957) and Piaget's (1930/1969) agent-oriented accounts.

5.3. Correlations and the Sense of Force

At the beginning of the paper, we noted that there is a difference between correlations that are causal and those that are noncausal. If forces are absolutely necessary for causation, then maybe the ingredient that makes a correlation causal is whether it involves force. This possibility was examined in an experiment involving abstract correlations. Specifically, participants ($N = 50$) saw animations showing two circles, like those shown in the last panel in Figure 5.1. The circle on the left can be viewed as the "cause" (C) and the circle on the right as the "effect" (E). In half of the trials, the cause turned solid, followed a few moments later by the effect turning solid. In 1/5 of the trials, the cause did not turn solid, but the effect did. Finally, in the remaining trials, neither circle turned solid. Based on these frequencies, the probability of the effect given the cause, $P(E|C)$, equaled 1, and the probability of the effect given the absence of the cause, $P(E|\neg C)$, equaled 0.4. These probabilities entail that the probability of the effect given the cause is greater than the probability of the effect in the absence of a cause, that is, $P(E|C) > P(E|\neg C)$; thus, the probabilities entail that the effect correlated positively with the cause.

As in the previous experiments, trials consisted of sets of four animations. However, in the current experiment, participants indicated only whether they felt a force. One other difference from the previous experiments is that at the end of the experiment, participants were asked several questions. First, they were asked whether it seemed that the circle on the left sometimes caused the circle on the right to change. Participants were also asked to estimate the percentage of times the circle on the right changed when the circle on the left changed, thus providing an estimate of $P(E|C)$, as well as the percentage of times the circle on the right changed when the circle on the left did NOT change, providing an estimate of $P(E|\neg C)$. Higher estimates for $P(E|C)$ than for $P(E|\neg C)$ would imply that participants noticed the correlation between the cause and effect. In sum, the questions at the end of the experiment allowed us to determine whether participants noticed a correlation and whether they felt the correlation was causal. In addition, because we also measured participants' responsiveness to forces, we could examine whether their judgments of causation were associated with their responsiveness to forces.

In Experiment 6, there were two conditions. Half of the participants were given mechanism information. Specifically, they were told that “The light on the left is linked to the one on the right through a long sequence of circuits.” The remaining participants were simply told that they would see a series of animations.

One of the main predictions was that people would be sensitive to the correlational structure of the materials in both the mechanism and no-mechanism conditions. A second main prediction was that people would be more likely to say that the first circle caused the second circle to change in the mechanism condition than in the no-mechanism condition. This prediction was based on pilot research showing that if people are given extremely sparse materials, they often fail to consider the possibility of a causal relationship unless they are given a cover story suggesting the existence of causal relation. The experiment was designed, then, to create a situation in which two groups of people encoded the same correlation, but only one of the groups would view the correlation as causal. The last main prediction concerned participants’ responsiveness to the forces generated by the haptic controller. If force is necessarily a part of the notion of causation, we should find greater responsiveness to forces in the mechanism condition than in the no-mechanism condition.

The results were as predicted. Firstly, as expected, participants noticed the correlation between the cause and effect circles in both conditions and to the same degree in both conditions. In support of this observation, the probability of the effect given the cause, $P(E|C)$, was significantly greater than the probability of the effect in the absence of the cause, $P(E|\neg C)$, in both the mechanism, $t(24) = 3.09, p < 0.05$, and no-mechanism conditions, $t(24) = 3.69, p < 0.01$, and there was no evidence that the difference in the probability estimates in the mechanism condition ($D = 22.5$) differed from the difference in probability estimates in the no-mechanism condition ($D = 22.4$), $F(1,48) = 0.026$.

The second major prediction was also borne out, in that participants endorsed the statement that the first circle seemed to cause the second circle to change more often in the mechanism condition ($M = 76\%$) condition than in the no-mechanism condition ($M = 44\%$), $t(48) = 2.39, p < 0.05$. Hence, participants picked up on the correlation between the circles in both the mechanism and no-mechanism conditions, but in terms of causation, the modal response in the mechanism condition was that it was causal, whereas in the no-mechanism condition it was noncausal. The question, then, is what makes a correlation seem causal? Turning to our last major prediction, as

expected, participants responded faster to a force in the mechanism condition ($D = -22.9$ ms), $t(24) = 4.43$, $p < 0.001$, than in the no-mechanism condition, ($D = -11.28$ ms), $t(24) = 1.88$, $p = 0.072$. Interestingly, dividing the participants in the no-mechanism condition according to whether they viewed the two circles as causally connected reveals that sensitivity to force was much greater in those who reported feeling there was a causal connection ($D = -21.26$ ms) than in those who did not feel there was a causal connection ($D = -2.8$). This implies that the marginally significant effect of force in the no-mechanism condition was driven completely by those who reported feeling there was a causal connection between the circles. The results paint a clear picture: the difference between correlations that are viewed as causal and those that are not viewed as causal is the feeling of force.



6. THE FEELING OF CAUSATION IN THE ABSENCE OF MECHANISM: A DUAL PROCESS APPROACH

The results discussed above suggest that people can feel a sense of force in the absence of a clear understanding of mechanism. This is surprising finding because if the underlying mechanism is not known, there is some question why forces should be felt. One possibility is that the perception of forces from visual materials is initially tied, developmentally speaking, to situations in which the mechanism is clearly present, that is, to situations in which physical contact, or a chain of physical contacts is present. With experience, people may learn to associate the existence of forces with only a subset of the properties in such scenes, like temporal contiguity. Such properties might develop into cues or heuristics for inferring the presence of forces. Eventually, such cues might lead people to feel forces even when knowledge of the underlying mechanism is not available. Moreover, to the extent that force impressions are based on heuristics, it should be expected that people will sometimes sense forces—and have a related feeling of causation—even when such forces are not actually present. By definition, heuristics are not infallible; they merely serve as rough guides to the existence of certain features of the environment. This proposal is supported by the existence of causal illusions, such as those that sometimes accompany electrical blackouts.

6.1. Explaining Causal Illusions in a Dual-process Framework

Back in July of 1977, a lightning strike hit the Buchanan South electrical substation on the Hudson River, tripping two circuit breakers in the

northern suburbs of Westchester County, New York. The event triggered a series of breakdowns that within 15 min left New York City in darkness. Interviews with “survivors” (Sparrow, 1999) indicated there was keen interest in the causes behind this event. Some attributed the blackout to equipment failure and others to unidentified flying object (UFOs) or to a Soviet invasion. Especially interesting, some felt, at least momentarily, that the blackout was caused by their own actions, like the opera singer who touched a door knob at the exact instant the lights turned off, or the child who accidentally hit a ceiling light fixture with her paddle ball, again at the exact moment everything went dark. As one of the 1997 survivors exclaimed after plugging in a toaster, “I blew out the whole neighborhood!” (Sparrow, 1999). Of course, touching a knob, hitting a ceiling light or plugging in an appliance cannot cause a massive blackout, but when the conditions are exactly right, a feeling of causation may emerge nonetheless.

Interestingly, in situations like blackouts, people may experience feelings of causation while at the same time knowing that such feelings are unwarranted. The fact that people can have conflicting opinions about the existence of a causal relationship is consistent with the idea that causal understanding might be based on two processes: an intuitive process that is fast and automatic and a reflective process that is slow and strategic. The intuitive process would be based on perceptual heuristics that give rise to a general sense of force and causation. The slow process would be one that depends on a careful analysis of the situation to determine whether there exists a mechanism for connecting the candidate cause and effect. The slow process would not necessarily involve checking every possible link in a chain. It might be satisfied with knowledge of a physical connection. However, in order to know whether causation is actually present, at the level of certainty required in science, for example, a detailed analysis of the mechanism would require knowledge not just of a physical connection, but also the forces or energies that are enabled by the physical connection.

The distinction between intuitive and reflective processes is not new. It maps directly unto a prominent distinction made in the perception, reasoning, and social cognition literature. According to dual-processing theories, System 1 processing refers to computations that are implicit, unconscious, and heuristic, while System 2 processing refers to computations that are explicit, analytic, and rule-based (Evans, 2008; Kahneman, 2003; Sloman, 1996; Stanovich & Toplak, 2012; Stanovich & West, 2000). Moreover, the distinction between intuitive and reflective processes aligns well with the distinction often made (usually implicitly) in the causation literature

between perceived and judged causation (Schlottmann & Shanks, 1992). Perceived causality is causality acquired directly from perceptual experience without aid from background knowledge (Leslie, 1988; Michotte, 1963; Rips, 2011; White, 2006). In contrast, judged causality is causality learned through elaboration and inference (Michotte, 1963; White, 2006; Schlottmann et al., 2006). While a distinction between perceived and judged causation has been noted in the literature, its empirical foundations have yet to be firmly established.

The idea that causal understanding might be based on two kinds of process, intuitive and reflective, offers a potential solution to the problem of how people might represent causal relations in terms of forces in the absence of knowledge of the underlying mechanism. Specifically, in particular situations, various properties of the environment might serve as cues to the presence of forces. One cue, in particular, would be temporal contiguity between an agent's actions and a particular effect. In line with dual-processing theories, impressions based on such cues might last only a moment. Moreover, the initial causal impressions should be especially sensitive to the properties of the cues. We would expect, then, that temporal contiguity might have a larger impact on people's momentary causal judgments than on their reflective judgments. This prediction was tested in the following experiment.

6.2. An Initial Test of the Dual Process View of Causal Understanding

In this experiment, participants ($N = 104$) saw a single animation depicting two main events: a person hitting a fire hydrant with a stick and an illuminated town going dark. Snapshots of two moments of these events are shown in Figure 5.3. The left panel of this figure shows a lit-up town in the background and, in the foreground, a person preparing to hit a fire hydrant with a stick. The panel on the right shows the situation at a later point in the animation, with the town completely dark. After watching the animation, participants were asked to imagine that they were the person in the scene and answer two questions: "Would you feel for a moment that the striking of the fire hydrant caused the lights to go out?" and "Would you ultimately conclude that the striking of the fire hydrant caused the lights to go out?" Participants recorded their answer on an eight-point scale in which 0 equaled "Definitely not" and 7 equaled "Definitely yes". For the sake of discussion, we report the results simply in terms of the proportion of times people reported "yes," that is, the proportion of times they gave a rating of 4 or higher.

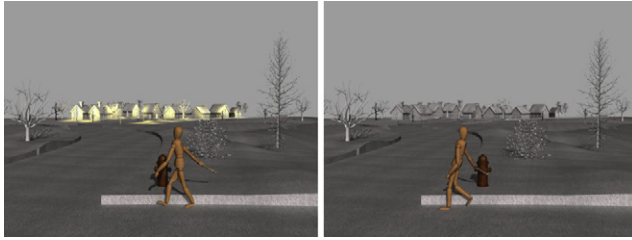


Figure 5.3 In the picture on the left, a person prepares to hit a fire hydrant with a stick, while in the picture on the right, the fire hydrant has been hit and the lights in the town have turned off. For color version of this figure, the reader is referred to the online version of this book.

We predicted that our blackout scenario would produce some of the same reactions reported during the 1977 New York City blackout, that is, our blackout scenario would give rise to a causal illusion. Specifically, we predicted that if the town went dark immediately after the person hit the fire hydrant, this would give rise to a causal impression in which people would *momentarily feel* as if the person in the scene caused the blackout. However, if people were asked to give a causal judgment by having them decide whether they would ultimately conclude that the person caused the town to go dark, we predicted that they would say “no”, that is, that they would recognize that the feeling of causation was not a true representation of what had happened.

In addition to testing this main prediction, the experiment was designed to test one further prediction, specifically, that temporal contiguity would have a larger effect on people’s initial causal impressions than on their later causal judgments. The prediction is based in part on work showing that temporal contiguity has an effect on people’s causal impressions of collisions events (Morris & Peng, 1994; Oakes & Kannass, 1999; Schlottmann & Anderson, 1993; Schlottmann & Shanks, 1992) than on their causal judgments of correlations (Schlottmann & Shanks, 1992). Such an effect is important because it would help establish the existence of two kinds of processes underlying causal understanding.

To test whether temporal contiguity had a larger effect on initial impressions than on more reflective judgments, participants were randomly assigned to one of four conditions (all $N = 26$). In the first condition, the town went dark immediately after the person hit the fire hydrant with the stick. In the second and third conditions, the town lights turned off one and 2 s respectively after the person hit the fire hydrant. Finally, in a

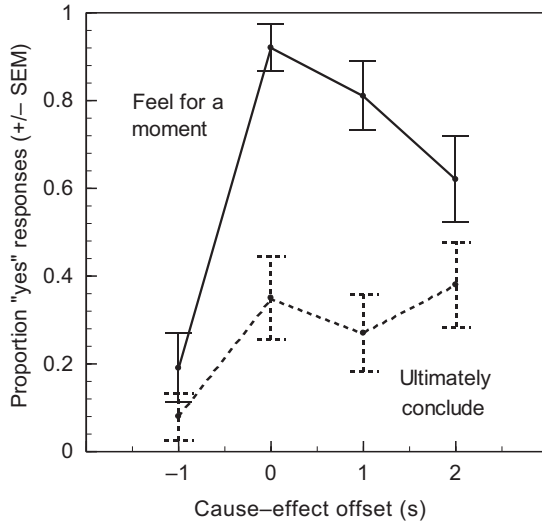


Figure 5.4 Proportion of “yes” responses to questions asking whether participants “felt” and “ultimately concluded” that the blackout event depicted causation for four different cause–effect offsets: the effect occurred before the cause by 1 s (–1), the effect occurred at exactly the same time as the cause (0), the effect occurred 1 s after the cause (1), and the effect occurred 2 s after the cause. The error bars represent standard errors of the mean.

fourth condition, the town lights turned off 1 s *before* the person hit the fire hydrant. In this fourth condition, the “effect” occurred before the cause, so it was predicted that participants’ impressions and judgments of causation should be very low. This condition was included to serve as a baseline condition for interpreting the results in the other conditions.

As shown in Figure 5.4, the results were in line with our predictions. Perhaps most strikingly, participants’ causal impressions patterned very differently from their causal judgments. For example, when the town darkened immediately after being hit (i.e. cause–effect offset = 0), 92% of the participants reported feeling a sense of causation, but only 35% reported that they would ultimately conclude that the striking of the fire hydrant caused the town to go dark. Another key difference between participants’ causal impressions and causal judgments was that temporal contiguity had a large impact on their causal impressions but not on their causal judgments. As the delay between the striking and the town darkening increased, participants’ causal impressions decreased from 92% in the 0 delay condition to 62% in the 2-s delay condition, a significant difference. In contrast, participants’ causal judgments stayed essentially the same across these two conditions, with 35%

yeses in the 0 delay condition and 38% yeses in the 2-s delay condition. As expected, participants' causal impressions and judgments were very low when the "effect" occurred before the "cause." Participants' causal judgments were lower in this condition than in the 0 and 2-s delay conditions. This finding implies that the absence of change in participants' causal judgment ratings across the different time offsets cannot be explained as due to their ratings being at floor. Moreover, they imply that participants may have found ways to explain how hitting a fire hydrant might result in a blackout.

Several conclusions can be drawn from these results. First, the results provide support for a distinction between intuitive and reflective processes in causal understanding. Causal impressions emerge spontaneously with little connection to prior causal knowledge. Causal judgments, in contrast, are based on more strategic thought involving prior knowledge. A second conclusion is that people may feel a strong impression of causation even when the generative process is unclear. As noted earlier, this phenomenon may explain how a force-based view of causation is possible when knowledge of the underlying mechanism is not available. Cues in the environment may trigger the impressions of forces that give rise to the impression of causation, even when the impression of causation is unwarranted. Thirdly, the results indicate that causal impressions are not limited to the kinds of scenarios originally studied by Michotte (1946/1963). In the launching event, both temporal and spatial contiguity were found to be important for the impression of causation (Schlottmann & Anderson, 1993; Schlottmann & Shanks, 1992; Scholl & Tremoulet, 2000; White, 2011). In the blackout scenario, the spatial arrangement of the A and B objects was quite different from Michotte's materials (1946/1963), yet an impression of causation was found. The results demonstrate that the causal impression can emerge from a much wider range of events than has generally been assumed.



7. POTENTIAL PROBLEMS FOR A FORCE-BASED VIEW OF CAUSATION

The evidence from a variety of sources offers support for the proposal that causal understanding is based on the feeling of force. There is, however, at least one study that raises a possible challenge to this proposal. In particular, recent findings from White (2011) have been interpreted as showing that impressions of causation and of force are independent of each other. This conclusion was based on the finding that participants' ratings of causation and force sometimes patterned differently, but not always. In the

first of several experiments, White (2011) found that impressions of force decreased with increases in time delay, just as Michotte (1963) found with impressions of causation. Thus, with respect to time, White (2011) found that impressions of force greatly resembled impressions of causation. However, in White's (2011) Experiment 4, he found that ratings of causation and force differed. Specifically, he found that ratings of force were weakly affected by differences in gap size and type of intermediary, while ratings of causation were greatly affected by these properties. These inconsistencies led White (2011) to conclude that the impression of causation is not based on the same process that underlies the impression of force.

It should be emphasized that White's (2011) experiment was probably the first to directly measure both impressions of causation and force, making it an important contribution to the field. It should also be noted, however, that the instructions used in White's (2011) Experiment 4 were not parallel across the conditions. Participants in the causation condition were told that events with gaps *were not* consistent with causation in real life, while participants in the force condition were told the opposite, namely, that gaps *were* consistent with forces in the real world because forces can be transmitted across gaps. Given these differences, it is certainly possible that the differences in participants' impressions of causation and their impressions of force were due to the differences in the instructions.

But even if the rating differences were not due to differences in the instructions, there are several other reasons why White's (2011) results do not rule out causation being based on force. As noted in the preceding section, ratings of causation may be based on either intuitive or reflective processes, or both. If ratings of force differ from ratings of causation, the difference could be due to the kind of causal process being tapped during the rating. As shown in the research on blackouts, the same event can be associated with two very different ratings of causation. One final reason why White's (2011) results need not rule out a causation-force connection is because the difference between the two kinds of ratings could be due to the way forces were measured. In White (2011), forces were measured using explicit verbal ratings, whereas in the haptic experiments described earlier, forces were measured using implicit reaction times. It remains an open question whether these two measures of force capture the same impression. To the extent that they do not, the two measures might lead to very different conclusions about the relationship between causation and force. Given all of these concerns, the conclusion that causation is not based on force seems, at best, premature. In White's (2012a, 2012b) most recent writings, it seems he would agree.



8. CONCLUSIONS

The idea that causation might be based on forces is one of the oldest and longest-lasting theories of causation. Major criticisms of this theory have focused on issue of learnability and whether such a theory can be applied to abstract causal relations. Recent work on the theory suggests that these two challenges can be met. With regards to learnability, the claim was that if causal relations were based on forces, then they could not be learned because forces are not part of people's sensory experience. While it is true that forces cannot be seen, this does not entail that they are absent from people's sensory experience because forces can definitely be felt, and because they can be felt, they can ultimately be learned. The sense of touch, proprioception, and balance provide a foundation for the notion of force, and then by extension, the concept of causation. The feeling of causation they provide may be extended to causal relations perceived through the visual modality. For causal relations perceived through vision, the sense of touch provides a sensory experience that unifies the vast range of visual patterns we associate with causation. With regards to the applicability of force-based theories to abstract causation, the experiments described in this paper show that a force-based approach can be extended to abstract causation. With time, the perception of forces becomes associated with cues to causation that allow for impressions of causation in the absence of knowledge of the underlying mechanism. Inferring forces from cues will sometimes lead to causal illusions, such as the illusion that a person caused an entire city to darken simply by hitting a fire hydrant. More often than not, however, these cues will lead people to infer causal relations that are actually present. According to this proposal, causal understanding remains physical in its phenomenology, even as its ontology is extended from the physical to the abstract. The great leap forward regarding causal understanding is in our ability to use secondary sensations—sensations experienced through another modality—to create conceptual structure.

REFERENCES

- Ahn, W., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology*, 31, 82–123.
- Ahn, W., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil, & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–225). Cambridge, MA: MIT Press.
- Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, 54, 299–352.

- Aristotle. (1999). *Aristotle Physics: Book VIII (D. Graham, Trans.)*. New York: Oxford University Press.
- Aristotle. (2011). *The Eudemean Ethics (A. Kenny, Trans.)*. New York: Oxford University press.
- Aronson, J. L. (1971). On the grammar of "cause". *Synthese*, 22, 414–430.
- Bigelow, J., Ellis, B., & Pargetter, R. (1988). Forces. *Philosophy of Science*, 55, 614–630.
- Bigelow, J., & Pargetter, R. (1990). Metaphysics of causation. *Erkenntnis*, 33, 89–119.
- Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. Friedman (Ed.), *The developmental psychology of time* (pp. 209–255). London: Academic Press.
- Blakemore, S. J., Bristow, D., Bird, G., Frith, C., & Ward, J. (2005). Somatosensory activations during the observation of touch and a case of vision–touch synaesthesia. *Brain*, 128, 1571–1583.
- Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review*, 104, 367–405.
- Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, 40, 83–120.
- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, 99, 365–382.
- Condit, M. A., Gandolfo, F., & Mussa-Ivaldi, F. A. (1997). The motor system does not learn the dynamics of the arm by rote memorization of past experience. *Journal of Neurophysiology*, 78, 554–560.
- Davidson, P. R., & Wolpert, D. M. (2004). Internal models underlying grasp can be additively combined. *Experimental Brain Research*, 155, 334–340.
- Dowe, P. (2000). *Physical causation*. Cambridge, England: Cambridge University Press.
- Ebisch, S. J. H., Perrucci, M. G., Ferretti, A., Del Gratta, C., Romani, G. L., & Gallese, V. (2008). The sense of touch: embodied simulation in a visuotactile mirroring mechanism for observed animate or inanimate touch. *Journal of Cognitive Neuroscience*, 20, 1611–1623.
- Ellis, R., & Lederman, S. J. (1999). The material-weight illusion revisited. *Perception & Psychophysics*, 61, 1564–1576.
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278.
- Fair, D. (1979). Causation and the flow of energy. *Erkenntnis*, 14, 219–250.
- Fales, E. (1990). *Causation and universals*. London: Routledge.
- Flanagan, J., & Beltzner, M. A. (2000). Independence of perceptual and sensorimotor predictions in the size-weight illusion. *Nature Neuroscience*, 3, 737–741.
- Gordon, A. M., Forssberg, H., Johansson, R. S., & Westling, G. (1991). Visual size cues in the programming of manipulative forces during precision grip. *Experimental Brain Research*, 83, 477–482.
- Hamilton, A., Joyce, D. W., Flanagan, J. R., Frith, C. D., & Wopert, D. M. (2007). Kinematic cues in perceptual weight judgment and their origins in box lifting. *Psychological Research*, 71, 13–21.
- Harré, R., & Madden, E. H. (1975). *Causal powers: a theory of natural necessity*. Oxford: Blackwell.
- Heidegger, M. (1981/1995). *Aristotle's metaphysics 1–3: On the essence and actuality of force (translated by W. Brogan & P. Warnek)*. : Indiana University Press. (Original published in 1981).
- Hinder, M. R., & Milner, T. E. (2003). The case for an internal dynamics model versus equilibrium point control in human movement. *Journal of Physiology*, 549, 953–963.
- Hubbard, T. L., & Ruppel, S. E. (2002). A possible role of naïve impetus in Michotte's "launching effect" evidence from representational momentum. *Visual Cognition*, 9, 153–176.
- Hume, D. (1748/1975). Enquiry concerning human understanding. In L. A. Selby-Bigge (Ed.), *Enquiries concerning human understanding and concerning the principles of morals* (3rd ed.). Oxford: Clarendon Press. revised by P. H. Nidditch.

- Imamizu, H., Uno, Y., & Kawato, M. (1995). Internal representations of the motor apparatus: implications from generalization in visuomotor learning. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1174–1198.
- Jammer, M. (1957). *Concepts of force*. Cambridge, MA: Harvard University Press.
- Jones, L. A. (2000). Kinesthetic sensing. In *Proceedings of workshop on human and machine haptics* (pp. 1–10). Cambridge, MA: MIT Press.
- Jones, L. A., & Pateski, E. (2006). Contribution of tactile feedback from the hand to the perception of force. *Experimental Brain Research*, *168*, 298–302.
- Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist*, *58*, 697–720.
- Kawai, S., Henigman, F., MacKenzie, C. L., Kuang, A. B., & Faust, P. H. (2007). A reexamination of the size-weight illusion induced by visual size cues. *Experimental Brain Research*, *179*, 443–456.
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*, 718–727.
- Keyers, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature Reviews*, *11*, 417–428.
- Keyers, C., Wicker, B., Gazzola, V., Anton, J., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, *42*, 335–346.
- Kotovskiy, L., & Baillargeon, R. (1998). The development of calibration-based reasoning about collision events in young infants. *Cognition*, *67*, 311–351.
- Kotovskiy, L., & Baillargeon, R. (2000). Reasoning about collisions involving inert objects in 7.5-month-old infants. *Developmental Science*, *3*, 344–359.
- Leslie, A. M. (1988). The necessity of illusion: perception and thought in infancy. In L. Weskranz (Ed.), *Thought without language* (pp. 185–210). Oxford: Oxford Science Publications.
- Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. Hirschfeld, & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge, England: Cambridge University Press.
- Marmodoro, A. (2007). The union of cause and effect in Aristotle: physics III 3. *Oxford Studies in Ancient Philosophy*, *32*, 205–232.
- Meyer, K., Kaplan, J. T., Essex, R., Damasio, H., & Damasio, A. (2011). Seeing touch is correlated with content-specific activity in primary somatosensory cortex. *Cerebral Cortex*, *21*, 2113–2121.
- Michotte, A. E. (1963). *The perception of causality* (translated by T. R. Miles, & E. Miles). London: Methuen. (Original published in 1946).
- Milner, T. E., & Franklin, D. W. (2005). Impedance control and internal model use during the initial stage of adaptation to novel dynamics in humans. *Journal of Physiology*, *567*, 651–664.
- Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*, 949–971.
- Oakes, L. M., & Kannass, K. N. (1999). That's the way the ball bounces: Infants' and adults' perception of spatial and temporal contiguity in collisions involving bouncing balls. *Developmental Science*, *2*, 86–101.
- Ohta, K., & Laboissière, R. (2006). Underlying principles of trajectory formation for human movement in dynamical environments. *International Congress Series*, *1291*, 97–100.
- Ott, W. (2009). *Causation and laws of nature in early modern philosophy*. New York: Oxford University Press.
- O'Neill, E. (1993). Influxus physicus. In S. Nadler (Ed.), *Causation in early modern philosophy* (pp. 27–57). Penn State University Press.
- Panarese, A., & Edin, B. B. (2011). Human ability to discriminate direction of three-dimensional force stimuli applied to the finger pad. *Journal of Neurophysiology*, *105*, 541–547.

- Papaxanthis, C., Pozzo, T., & McIntyre, J. (2005). Kinematic and dynamic processes for the control of pointing movements in humans revealed by short-term exposure to micro-gravity. *Neuroscience*, *135*, 371–383.
- Piaget, J. (1930/1969). *The child's conception of physical causality*. (translated by Marjorie Gabain, 1969). Totowa, NJ: Littlefield, Adams & Co.
- Pihko, E., Nangini, C., Jousmäki, V., & Hari, R. (2010). Observing touch activates human primary somatosensory cortex. *European Journal of Neuroscience*, *31*, 1836–1846.
- Reid, T. (1788/2010). Sections from essays on the active powers of man. In T. Nadelhoffer, E. Nahmias, & S. Nichols (Eds.), *Moral psychology: Historical and contemporary readings* (pp. 246–250). Wes Sussex, United Kingdom: Blackwell Publishing Ltd.
- Reinkensmeyer, D. J., Emken, J. L., & Crammer, S. C. (2004). Robotics, motor learning, and neurological recovery. *Annual Review of Biomedical Engineering*, *6*, 497–525.
- Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science*, *6*, 77–97.
- Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, *112*, 585–615.
- Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science*, *61*, 297–312.
- Salmon, W. (1998). *Causality and explanation*. Oxford, England: Oxford University Press.
- Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica*, *123*, 144–165.
- Schlottmann, A., & Anderson, N. H. (1993). An information integration approach to phenomenal causality. *Memory and Cognition*, *21*, 785–801.
- Schlottmann, A., Ray, E., Mitchell, A., & Demetriou, N. (2006). Perceived physical and social causality in animated motions: spontaneous reports and ratings. *Acta Psychologica*, *123*, 112–143.
- Schlottmann, A., & Shanks, D. (1992). Evidence for a distance between judged and perceived causality. *Quarterly Journal of Experimental Psychology*, *44A*, 321–342.
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*, 299–309.
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, *47*(1, Serial No. 194), 1–51.
- Schulz, L., Kushnir, T., & Gopnik, A. (2007). Learning from doing: intervention and causal inference. In A. Gopnick, & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 67–85). Oxford: Oxford University Press.
- Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, *14*, 3208–3224.
- Shim, J., & Carlton, L. G. (1997). Perception of kinematic characteristics in the motion of lifted weight. *Journal of Motor Behavior*, *29*, 131–146.
- Slooman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
- Sparrow, J. T. (1999). *The blackout history project: A virtual machine that would go of itself*. Retrieved from <http://sloan.stanford.edu/SloanConference/papers/sparrow.html>.
- Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). Oxford: Clarendon Press.
- Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of type 1 and type 2 processing. *Mind and Society*, *11*, 3–13.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral Brain Sciences*, *23*, 645–726.
- Strawson, G. (1987). Realism and causation. *The Philosophical Quarterly*, *37*, 253–277.
- Straube, B., & Chatterjee, A. (2010). Space and time in perceptual causality. *Frontiers in Human Neuroscience*, *4*, 1–10.

- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, 12, 49–100.
- Truman, N. E. (1904). *Maine de Biran's philosophy of will*. New York: The MacMillan Company.
- Walker, P., Francis, B. J., & Walker, L. (2010). The brightness-weight illusion. *Experimental Psychology*, 57, 462–469.
- Wheat, H. E., Salo, L. M., & Goodwin, A. W. (2004). Human ability to scale and discriminate forces typical of those occurring during grasp and manipulation. *The Journal of Neuroscience*, 24, 3394–3401.
- White, P. A. (1999). Toward a causal realist account of causal understanding. *American Journal of Psychology*, 112, 605–642.
- White, P. A. (2006). The role of activity in visual impressions of causality. *Acta Psychologica*, 123, 166–185.
- White, P. A. (2009). Perception of forces exerted by objects in collision events. *Psychological Review*, 116, 580–601.
- White, P. A. (2011). Visual impressions of force exerted by one object on another when the objects do not come into contact. *Visual Cognition*, 19, 340–366.
- White, P. A. (2012a). The experience of force: the role of haptic experience of forces in visual perception of object motion and interactions, mental simulation, and motion-related judgments. *Psychological Bulletin*, 138, 589–615.
- White, P. A. (2012b). Visual impressions of causality: effects of manipulating the direction of the target object's motion in a collision event. *Visual Cognition*, 20, 121–142.
- Witt, C. (2008). Aristotelian powers. In Ruth Groff (Ed.), *Revitalizing causality: Realism about causality in philosophy and social science*.
- Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, 136, 82–111.
- Wolff, P., Barbey, A. K., & Hausknecht, M. (2010). For want of a nail: how absences cause events. *Journal of Experimental Psychology: General*, 139, 191–221.
- Wolff, P., Ritter, S., & Holmes, K. Causation, forces, and the sense of touch, in preparation.
- Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology*, 47, 276–332.
- Wolff, P., & Zettergren, M. (2002). A vector model of causal meaning. In W. D. Gray, & C. D. Schunn (Eds.), *Proceedings of the twenty-fourth annual conference of the cognitive science society* (pp. 944–949). Mahwah, NJ: Erlbaum.
- Woodward, J. (2007). Interventionist theories of causation in psychological perspective. In A. Gopnik, & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 19–36). Oxford: Oxford University Press.



Categorization as Causal Explanation: Discounting and Augmenting in a Bayesian Framework

Daniel M. Oppenheimer^{*,†,1}, Joshua B. Tenenbaum[§], Tevye R. Krynski[§]

^{*}Department of Psychology, Princeton University, Princeton, NJ, USA

[†]Anderson School of Management, University of California, Los Angeles, CA, USA

[§]Department of Brain and Cognitive Sciences, M.I.T., Cambridge, MA, USA

¹Corresponding author: E-mail: daniel.oppenheimer@anderson.ucla.edu

Contents

1. Introductory Comments	204
2. Previous Accounts of Categorization	206
2.1. Feature-Based Accounts of Categorization	206
2.1.1. <i>Prototype Models</i>	207
2.1.2. <i>Exemplar Models</i>	208
2.1.3. <i>Mixture Model</i>	209
2.2. Models Based on Causal Reasoning	210
2.2.1. <i>Murphy and Medin</i>	210
2.2.2. <i>Causal Determinants of Feature Centrality</i>	211
2.2.3. <i>Causal Model Theory</i>	212
3. The Present Model	212
3.1. Impetus for the Present Model	212
3.2. A Causal Bayesian Network Framework for Categorization	213
4. Tests of the Predictions	217
4.1. Design of Tests	217
4.2. Predictions of the Model	219
4.3. Do Category-Irrelevant Contextual Features Affect Categorization	219
4.4. Can Category-Irrelevant Contextual Features be Preventative as well as Generative?	221
5. Placing the Present Account in the Literature	223
5.1. Other Accounts of Discounting	223
5.2. Relating the Bayesian Network Model to Other Models of Causally Based Categorization	225
5.3. Caveats	227
6. Conclusion	228

Abstract

Most empirical studies and formal models of theory-based categorization have focused on how features within a concept are causally related. In this paper, we explore the possibility that concepts and their features should be thought of in a more holistic manner, as embedded in a much larger and context-dependent web of causal connections which can impact categorization. We present a Bayesian model designed to integrate models of categorization, from the classic feature-based view, through concept-constrained causal-model views, to the more holistic causal reasoning implied by Murphy and Medin's (1985) paper. A set of empirical findings demonstrating causal discounting (Experiment 1) and augmenting (Experiment 2) provide evidence that concept-irrelevant causal knowledge is taken into account in categorization judgments. The implications for this model in terms of broader research on categorization are discussed.



1. INTRODUCTORY COMMENTS

Imagine that while walking down the street you hear a noise coming from a nearby ditch. You can see an animal rooting around, but you cannot quite make out whether it's a skunk or a raccoon. Judging by the awful smell in the air you decide it must be a skunk. However, after walking on a little further you notice that the ditch is actually a sewage duct. In a case like this, you might reason as follows: I had thought the animal was a skunk because of the terrible smell, but I guess the smell may have been due to the sewer instead. In which case, maybe the thing I saw could have been a raccoon after all.

Categorization does not take place in a vacuum. In the real world, there are a vast number of contextual cues that could play a role in categorizing an object, person, or situation. Contextual features can play a role in whether or not a particular feature is a valid cue to category membership. Although the acceleration of a car might normally be very useful in determining whether it is a Porsche, if the car is going down a particularly steep hill, then the rapid acceleration rate may no longer be diagnostic. Similarly, a slow acceleration rate might not be particularly informative when the car is driving through a school zone, being trailed by a police car, or is stuck in bumper-to-bumper traffic. The goal of this paper is to give a computational account of these sorts of context-sensitive categorizations in terms of a Bayesian framework for causal explanation and to present evidence for this model across multiple real-world domains.

A long tradition of formal models have examined how the features of a given entity can be compared with representations of categories in memory

to determine the entity's category membership (e.g., Anderson, 1991; Fried & Holyoak, 1984; Gluck & Bower, 1998; Nosofsky, 1986; Medin & Smith, 1981). This notion of categorization as a bottom-up, feature-matching process was criticized by Murphy and Medin (1985), who argued that categorization often is seen better as a top-down process of explanation, drawing on people's intuitive causal theories of physical, biological, and social domains. Their example of categorizing a man who jumps fully clothed into a swimming pool at a party as drunk is perhaps the prototypical example of categorization as causal explanation.

Partly inspired by Murphy and Medin's (1985) critique, more recent models have begun to focus on the role of causal reasoning in categorization (e.g., Ahn, 1999; Rehder, 2003; Waldmann, Holyoak, & Fratianne, 1995). These models treat knowledge about a category as a causal network of features associated with category members, in which entities are judged as good members of category "C" to the extent that their features were likely to have been produced from C's causal network. However, there are no formal models of categorization based on features that are not typically associated with a category (such as using "jumped into a swimming pool fully clothed" as a cue for being drunk; Murphy & Medin, 1985). In fact, although the logic described in the skunk example seems both natural and ubiquitous, there are presently no formal models that attempt to account for such reasoning. Ideally, a model of categorization should be able to take into account both the manner in which features are combined, and how those features can interact with context.

Mathematically formalized feature-based accounts of categorization have typically been viewed as incompatible with more intuitively framed notions of theory-driven categorization. In this paper, we seek to unify these views in a common framework. Using Bayesian networks (Pearl, 1988) to represent category knowledge, we model category judgments as Bayesian inference to the best explanation for a given set of features. Previous approaches to categorization, both feature-based and causal or theory-driven, can be shown to be special cases of this more general framework.

Our paper proceeds as follows: First, we review previous accounts of categorization, focusing on how each account can be modeled using the language of Bayesian networks. These Bayesian network treatments of previous models will in some cases appear nonstandard, but that will enable us to clarify exactly how our proposal formally relates to and extends these well-known accounts of categorization. Following this review, we present a single unifying framework that subsumes each of these individual Bayesian

network models. This presentation will also describe how contextual information not typically viewed as part of category knowledge can be represented in the framework and play a role in categorization. Finally, we will provide some preliminary evidence for the model, and discuss the implications of this approach for future research in categorization.



2. PREVIOUS ACCOUNTS OF CATEGORIZATION

2.1. Feature-Based Accounts of Categorization

Feature-based models typically assume that categories are associated with a distribution of feature values in a multidimensional space, and that items are categorized based on how well their feature values match with each potential category's characteristic feature distribution. Formally, categorization can be modeled as a probabilistic inference about how likely an item's particular feature values are to be observed for a random sample of that category. While feature-based models come in a variety of forms, the most prominent types are exemplar models (Medin & Schaffer, 1978; Nosofsky, 1986), prototype models (Fried & Holyoak, 1984), and mixture models (Anderson, 1991; Rosseel, 2002; Sanborn, Griffiths, & Navarro, 2006).

Each of the above accounts of categorization shares a common principle: the features of an entity are assumed to depend probabilistically on the entity's category membership, and nothing else. Anderson's (1991) mixture model expresses this principle explicitly, by specifying a distribution over feature values for each category. Bayesian networks (Pearl, 1988) are a tool specifically designed to capture such dependencies, and we can use them to better understand the dependency structure of the various feature-based models. Figure 6.1 depicts a version of Anderson's model as a Bayesian network in which a single variable, C , is the parent of each of the features

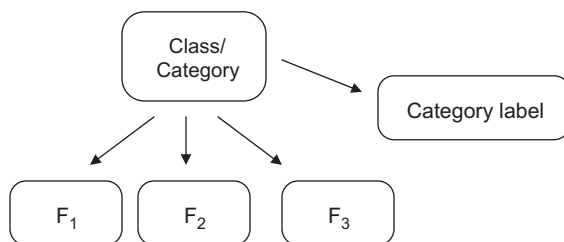


Figure 6.1 A depiction of a version of Anderson's model as a Bayesian network in which a single variable is the parent of each of the features. In a prototype model, the values of this variable can be interpreted as categories, but in other models, the values are more appropriately interpreted as abstract classes.

(Figure 6.1). In a prototype model, the alternative states of C can be interpreted as verbally labeled categories (Figure 6.2). In other models, the values of C are more appropriately interpreted as abstract classes of stimuli sharing the same expectations about likely category labels and other feature values, but not necessarily in one-to-one correspondence with the verbally labeled categories (Figure 6.3). Along with the model structure of Figure 6.1, a Bayesian network must specify a probability distribution over the values of the features for each value of C . Because the distribution of features depends on the value of C , members of the same class are likely to share features. Nosofsky (1991) showed that Anderson's mixture model can interpolate between prototype and exemplar views depending on the coupling parameter, which controls how likely a newly encountered item is to come from an existing class versus a new class. The Bayesian network of Figure 6.1 can also be used to interpolate between prototype and exemplar views, as we will show below. We will also discuss an extension of this model that can account for cases in which the features are dependent on each other, as described by Rehder (2003). See Danks (2007) for an alternative approach to unifying some of these same models of categorization in terms of Bayesian networks and probabilistic graphical models.

2.1.1. Prototype Models

Prototype models assume that for each category people retain in memory a single specific example (the prototype), and that category members in the world fall in a distribution around the prototype (Fried & Holyoak, 1984;

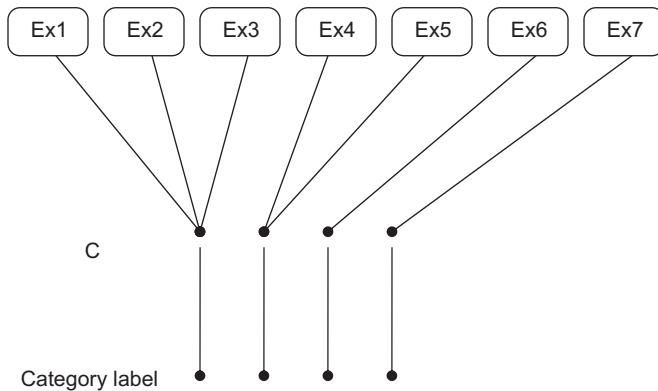


Figure 6.2 Learning a prototype model. Examples are classified by their similarity to existing classes (prototypes). New classes are formed for sufficiently distinct examples. Category labels have a one-to-one correspondence with classes.

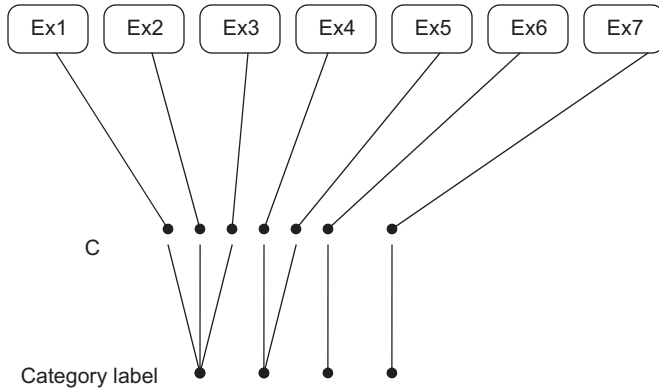


Figure 6.3 Learning an exemplar model. A new class (exemplar) is formed for each example. Category labels are associated with each class, according to the label observed for each example. New category judgments are made according to the similarity to previous examples.

Reed, 1972). Categorization is then a function of how similar the category's prototype is to the object to be categorized. Some prototype models, for example, assume that the similarity of an item to the prototype of a category is an exponentially decaying function of the psychological distance (or squared distance) between them (Nosofsky, 1987, 1992).

To model the prototype view with the Bayesian network model of Figure 6.1, the C variable would take on states representing different possible categories, and the feature values would depend probabilistically on which category the item is a member of (Figure 6.2). The similarity-based prototype models can be mathematically formalized by specifying a Gaussian distribution over feature values for each category, with the mean and variance of each distribution depending on the category. This method is formally identical to assessing the similarity of an item to the prototype using an exponentially decaying squared-distance function, as described above.

2.1.2. Exemplar Models

Exemplar models work under the assumption that individuals have a representation in memory of many (or all) members of each category. Individuals attempting to categorize an object compare that object to each of the stored exemplars. Categorization decisions are ultimately made based on which group of stored exemplars is more similar to the object in question. Perhaps, the most well-known exemplar model of categorization is Nosofsky's (1986) Generalized Context Model (GCM).

The GCM (Nosofsky, 1986) is an elaborated version of Medin and Schaffer's (1978) Context Model of categorization and has been particularly successful in explaining a wide range of empirical data concerning human categorization (Kruschke, 1992). GCM makes categorization predictions on the basis of the location of observed exemplars in a multidimensional stimulus space. In essence, the model works by examining the summed similarity between a novel exemplar and the stored exemplars of each possible category label. The model assumes that the similarity between the exemplars and the item to be categorized is a decaying function of the psychological distance between them. As in the prototype model, different distance functions are possible, but for simplicity and consistency with the treatment of prototype models above, we will focus on a squared Euclidean distance measure.

To model the exemplar view using the Bayesian network of Figure 6.1, the C variable would take on states representing each individual exemplar (Figure 6.3), and the feature values would be predicted by a Gaussian distribution around the values of the exemplar. Category judgments would be made by inferring the probability of the category label given the observed features, which essentially sums the probability that the item corresponds to each of the exemplars stored for the category label.

2.1.3. Mixture Model

In his rational analysis of categorization, Anderson (1991) proposed a model for how people might optimally learn categories from examples and compute category membership based on the learned categories. His account takes a *mixture model* to be the basis for representing the feature distributions characteristic of categories. The model generates items by mixing the distributions over features among the possible categories from which the item could have been generated. Anderson argues that categorization can be seen as a form of Bayesian inference, in which the probability of membership in a class is a function of the probability of having the observed features given membership in the class, as well as the prior probability of membership in the class.

As described in Nosofsky (1991), this model has the capability of interpolating between an extreme exemplar model, in which every exemplar forms its own class, and a prototype model, in which the classes correspond to categories and all items having the same category label are assigned to the same class. The coupling parameter of Anderson's (1991) model specifies the prior probability that a newly encountered item comes from an existing class, and controls whether the model acts more like a prototype or more

like an exemplar model. An exemplar model can be achieved by allowing several classes for each category label, and making the coupling parameter low enough that a new class is formed for each novel item encountered during learning (Figure 6.3). A prototype model can be achieved by making the coupling parameter high enough that new items are likely to come from an existing category and by specifying a one-to-one correspondence between classes and category labels (Figure 6.2). Interpolations between these two extremes can be achieved by varying the coupling parameter for forming a new class, such that new classes are only formed for newly encountered items that are sufficiently different from earlier items. The Bayesian network of Figure 6.1 was based on Anderson's model, and is in fact formally equivalent to it. The classes of Anderson's model correspond to the states of the C variable in the Bayesian network. The fact that this Bayesian network can be specialized to give networks corresponding to prototype or exemplar models is analogous to the insight that Anderson's rational model can interpolate between prototype and exemplar views of categorization.

2.2. Models Based on Causal Reasoning

Higher order theory-based models are an alternative to simple feature-based models. Theory-based models assume that, in addition to looking at low-level feature similarity, individuals take into account their theories about how features interact, and use causal reasoning in categorization judgments. Some of the most prominent of these views are Murphy and Medin's (1985) theory-based categorization, Ahn's (1998) feature centrality model, Waldmann *et al.*'s (1995) causal model theory, and the similarly named causal model theory of Rehder (2003).

2.2.1. Murphy and Medin

In their seminal paper, Murphy and Medin (1985) argued that exemplar and prototype models were incapable of using knowledge about the real world to constrain when and which features were used in categorization (for further elaboration of this point, see Murphy, 2002). They noted that feature-based models ignore the problem of what counts as a feature, and proposed that people use causal knowledge to solve this and related problems.

Murphy and Medin suggest several key elements that distinguish a theory-based approach from a similarity-based approach. For example, similarity approaches characteristically consider which attributes are correlated

with a given category, while the theory-based approach thinks about categorization as causal explanation of those feature correlations.¹

While [Murphy and Medin's \(1985\)](#) approach had a considerable impact on the field, it is best thought of as a conceptual framework rather than a theory or model. While it provides a useful way of thinking about the issue of categorization, it does not provide any formal predictions, or any specific mechanisms of how causal-based categorization would work. To remedy this, several researchers have begun to develop formal mathematical models to explain categorization as causal reasoning.

2.2.2. Causal Determinants of Feature Centrality

One particularly elegant instantiation of causal reasoning in categorization is the feature centrality approach ([Ahn, 1998, 1999](#); [Ahn, Kim, Lassaline, & Dennis, 2000](#)). This approach addresses the fact that although similarity-based accounts have parameters to account for differential feature weighting, those weights are usually assigned in an underspecified manner. Typical methods of assigning weights, such as perceptual salience and category validity, cannot account for empirical findings. For example, perceptual salience cannot explain why people are more likely to accept the notion of a square cantaloupe than a square basketball ([Medin & Shoben, 1988](#)). Squareness is equally invalid for both categories, and in both cases is perceptually salient; so why is it differentially weighted?

[Ahn \(1998\)](#) noted that causal knowledge could apply to this problem. Features that cause other features are more central to a category and will be weighted more heavily. The fact that a basketball is round causes it to bounce, whereas the fact that a cantaloupe is round does not cause any other features of a cantaloupe. In a series of clever studies, [Ahn et al. \(2000\)](#) repeatedly demonstrate that more causally important features play a larger role in category judgments than less causally important ones.

The feature centrality theory provides important insight into the particular mechanisms by which causal theories play a role in categorization judgments. However, the theory is designed to determine how features internal to a to-be-categorized object are weighted, not how environmental and contextual cues contribute to when the features should be used.

¹ It is important to distinguish the “theory” ([Gopnik & Meltzoff, 1997](#)) from “theory-based categorization” ([Murphy & Medin, 1985](#)). While these conceptual frameworks are frequently used interchangeably they are distinct in important ways. While [Medin and Murphy's \(1985\)](#) paper was entitled “The role of theories in conceptual coherence,” the paper was more about categorization as causal explanation, not theories in the way that most cognitive scientists and cognitive developmentalists typically use this term.

2.2.3. Causal Model Theory

Another approach to categorization using causal reasoning is Causal Model Theory (Rehder's, 2003; Waldmann *et al.*, 1995). As in Ahn's (1998) model of feature centrality, Rehder notes that some features cause others. Waldmann *et al.* (1995) and Rehder (2003) focused on the fact that differences in causal structure will lead to differences in the patterns of features. We shall focus our discussion to Rehder's version of Causal Model Theory in this paper because it is more explicitly linked to Bayesian networks, but the central line of reasoning applies to Waldmann's model as well.

According to Rehder's (2003) Causal Model Theory, individuals have not only a representation of which features are associated with which categories, but they also have a representation of a causal model that relates the features to one another, which predicts how those features should tend to co-occur. Rehder considers categorization to be a probabilistic judgment of how well that causal network explains the observed examples, *i.e.*, the probability that a set of observed features would be generated by a particular causal model.

Rehder's model is an important step forward in the operationalization of Murphy and Medin's (1985) conceptual framework of causal-based categorization. It can account not only for feature weights, but also for preferences in certain patterns of feature co-occurrence. However, Rehder focuses exclusively on causal relationships of features within a concept. As such, his model cannot account for how causal entities outside of the concept—such as environmental and contextual cues—are able to constrain or promote the utility of certain features in category judgment.



3. THE PRESENT MODEL

3.1. Impetus for the Present Model

Causal reasoning can play a large role in categorization judgments (Murphy & Medin, 1985). Most instantiations of this conceptual framework have dealt with a special case: explanations constructed with the knowledge of how features *within a concept* are causally related. These theories all assume that there is a unit of thought, consisting of a concept and its features, which can be carved off from the rest of knowledge. While such explanations offer a different approach to understanding categorization, by virtue of the fact that they neglect the importance of contextual knowledge, they are, in a sense, in the same camp as the feature-based view.

The implications of [Murphy and Medin's \(1985\)](#) insight are broader. One can think of concepts and their features as embedded in a broader web of causal relations in which *all* concepts and environmental/contextual cues may be causally connected. Intuitively, this appears to be the case in [Murphy and Medin's \(1985\)](#) classic example in which we categorize a fully clothed person who jumps into a pool at a party as “drunk”. It seems unlikely that people have a stable representation of the concept “drunk”, which includes causal relations between the features “has clothes on” and “jumps into pool”.

To fully take into account the richness of causal reasoning in categorization, it is important to look at how features of the concept, the context of the categorization judgment, and a variety of other external cues might interact. In the skunk example, while the feature of a nasty odor could be caused by the creature being a skunk, it could also be caused by the creature's walk through the sewers. Although sewer-wading is not likely to be part of our representation of the concept of skunk, we can still incorporate that information into our reasoning if we think about categorization as causal reasoning in a broader context. The goal of this paper is to propose and empirically validate a framework for integrating prior accounts of categorization from the classic feature-based view ([Holyoak & Fried, 1984](#); [Nosofsky, 1987](#)), through causal-model feature-based views ([Ahn, 1998](#); [Rehder, 2003](#); [Waldmann et al., 1995](#)) to the more holistic causal reasoning implied by [Murphy and Medin's \(1985\)](#) conceptual framework.

Next, we discuss our proposed framework, which is based on a causal interpretation of the Bayesian network model we have been referencing thus far, but integrated with additional causal knowledge. This framework, like Anderson's mixture model, is consistent with the exemplar view, the prototype view, and models that interpolate between them. For simplicity, however, we will present our framework in the context of the prototype model, where each category corresponds to a single class, though the principles hold for other types of models as well.

3.2. A Causal Bayesian Network Framework for Categorization

Inspired by the utility of Bayesian networks for modeling both feature-based and causal model approaches to categorization, we propose a Bayesian network framework that treats categorization as a form of causal inference. Our framework assumes that features associated with a category are, in some abstract way, *caused* by category membership. A second

assumption of our model is that the presence of features can also be caused by sources other than category membership. A bad odor can be caused by being a member of the skunk category, or by a romp through the sewers. The movie “Chicken Run” (Lord & Park, 2000) was based on the premise that chickens can be made to fly by use of a catapult (but not through being a chicken).²

Like previous causal theories of categorization, our approach can be modeled using Bayesian networks (Pearl, 2000). Because people prefer to learn and reason from causes to effects (Eddy, 1982; Einhorn & Hogarth, 1986; Waldmann & Holyoak, 1992; Waldmann *et al.*, 1995), Bayesian network models are often structured causally, making each effect probabilistically dependent on the presence of its causes. We advance a framework in which a variable representing category membership causally influences all category features. Each feature can also have other causal influences, either other features of the category (i.e., descendents of the category membership node, as in Rehder (2003)) or factors external to the category (i.e., neither parents nor descendents of the category membership node).

In the most general version of our framework, depicted in Figure 6.4, these external causes may interact with category membership (and with each other) in arbitrarily complex ways. Since people may be implicitly aware of the complex ways that the causes of each feature interact, it is important to model the interaction correctly when accounting for people’s judgments. To simplify matters, we designed experimental stimuli in which the causes interacted in a specific way: each cause independently contributed to the probability of the effect being present. We also simplified the model by avoiding the complexity of features causing other features in our stimuli. The model we use for our stimuli, a special case of our more general framework, is depicted in Figure 6.5.

²The notion that category membership is a common cause of an object’s features is consistent with the proposal that people are essentialists (Gelman, 2003; Locke, 1894/1975) when reasoning about categories. The basic premise of essentialism is that some unobservable quality (e.g., skunkness) is shared by all members of a category, and is responsible for their similarities. That is, the presence of an observable feature in each of a category’s members is caused by the category’s essence being present in each member (Putnam, 1977). While such an essence may not be a physical reality, it has been argued that it is a psychological reality in that people behave as though such an essence truly does exist (Medin & Ortony, 1989). There have been demonstrations that essentialist notions are held cross culturally (Atran, 1987) and in young children (Gelman & Welman, 1991), lending credence to the notion that this may be a fundamental characteristic of human reasoning. While our model is consistent with essentialism, it is also consistent with some nonessentialist views, and it is beyond the scope of this paper to address the ongoing debate between the two views.

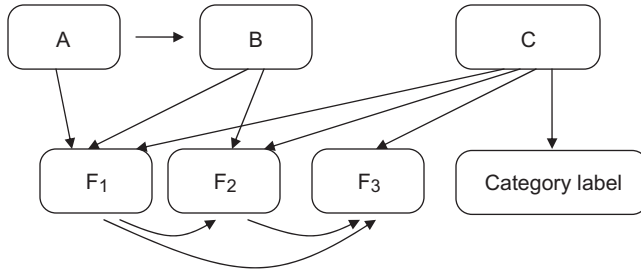


Figure 6.4 Our causal Bayesian network framework for categorization. The individual features depend on which class (C) the item is a member of, and can also depend on other features of the class that causally influence them (e.g., F1 causes F2, etc.), as well as causes external to the concept (A and B).

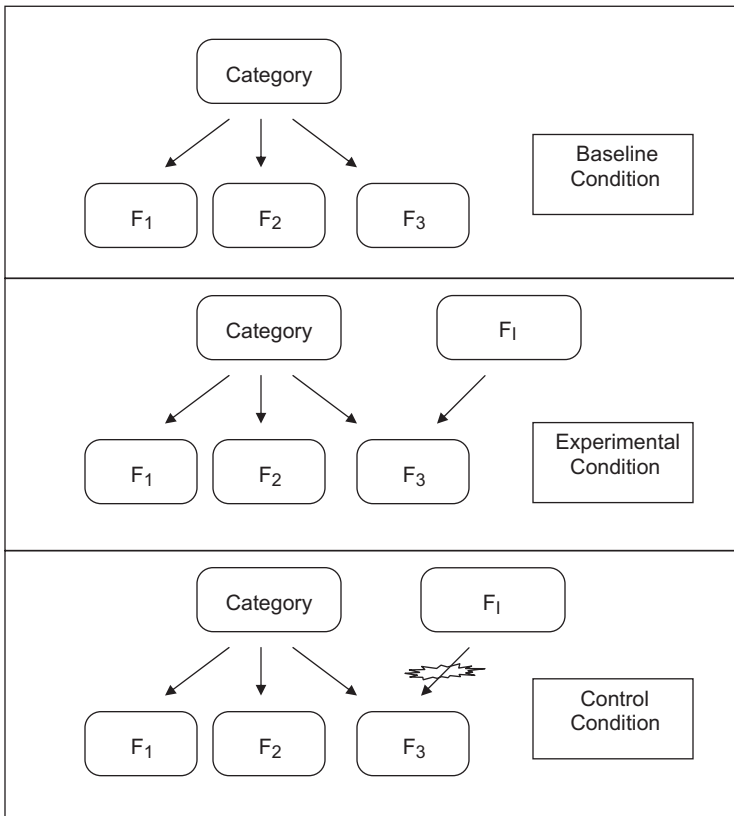


Figure 6.5 Graphical representation of experimental design. In the baseline condition, the category is described by three features. In the experimental condition, a category irrelevant feature (F_i) that is causally related to one of the diagnostic features is added. In the control condition, a category irrelevant feature that is not related to any other feature is added.

For our stimuli, we model the probability of a feature being present as a sigmoid function of the combined strength of its various causes, one of which is category membership (Figure 6.6). In this model, a cause only contributes to the probability of the effect if it is present, and causes can be either generative (positive strength) or preventative (negative strength):

$$P(F|c_1\dots c_N) = \left[1 + \exp \left(- \left[\sum_{i=1}^N s_i c_i \right] - L \right) \right]^{-1} \quad (1)$$

where s_i is the strength of the i th cause, c_i corresponds to whether the i th cause is present (i.e., 1 if present, or 0 if absent), and L is a leak term to account for the possibility of the effect occurring in the absence of identified causes. This function requires strengths to be assigned to the various causes; we assume these can be assigned based on prior knowledge that participants bring to the task. In addition, since a category judgment requires inferring the odds of a particular cause being present (category membership) from a known effect (feature), additional parameters are required to represent the prior probabilities (base rates) of each of the causes (including category membership) in the Bayesian network. Once we have specified these parameters, the probability that an object belongs to category C given that it has particular observed values for features $F1, F2, F3$, etc., can be computed by performing inference in the Bayesian network shown in Figure 6.5.

From a neurocomputational point of view, this sigmoidal model has a great deal of face validity for the types of scenarios we present. Imagine there was a “bad smell” neuron/detector in the brain that fired according to a sigmoidal function of the total amount of bad-smelling vapors coming into one’s nose. Skunks put out a certain amount of bad-smelling vapors, and so do animals that have been wading in the sewer. Dousing an animal with perfume counteracts some of the bad-smelling vapors. The total amount of bad-smelling vapors coming into one’s nose is then the sum total of these three terms.

Despite the validity of this sigmoidal model for some scenarios, we do not limit our framework to require this model, because it clearly does not apply to all domains. Three important requirements must be met for the sigmoidal model to be valid: (1) each cause must be capable of influencing the effect independently of the other causes, (2) the causes must not influence each other, and (3) when two of the causes are present at the same time, they must both be capable of influencing the effect.

Our framework allows for a great deal of flexibility in allowing one to incorporate knowledge external to a category into one's categorization judgments. It is also worth noting that in situations when external causes of the features are not explicitly represented, the sigmoidal model reduces to a basic prototype model, i.e., a simple feature-based account of categorization with features weighted according to the causal strength of category membership to generate them. But in its full generality, our framework is compatible with both Anderson's mixture model and Rehder's model with inter-feature causation.

Our framework differs from other accounts of categorization as causal reasoning in that it allows one to take causal knowledge external to traditional category knowledge into account. In particular, the presence of a generative cause for a feature aside from category membership should reduce one's confidence that an object with that feature is truly a member of the category. This notion of causal discounting (Kelley, 1972) has intuitive appeal. A feature present in the absence of any "external" causes (causes other than category membership) can only be explained through category membership. However, if alternate generative causes exist, then those could account for the existence of the feature, and one need not infer that the object is a member of the category. Our model strongly predicts discounting of category membership when generative external causes are present. The model also predicts augmenting when inhibitory external causes are present. If external causes work against the presence of a feature, then observation of that feature should lead one to infer that the cause of the feature must be category membership. If a car accelerates quickly despite driving uphill, it must have a really powerful engine, which makes it more likely to be a sports car.



4. TESTS OF THE PREDICTIONS

4.1. Design of Tests

To test these predictions, several experiments were run, with the following basic design: In a baseline condition, participants are told about an animal that has three features, F_1 , F_2 , and F_3 . These features are diagnostic of a category C . Participants are then asked about their confidence that the animal is a member of C . In the experimental condition an additional feature, F_E , is added to the original set of baseline features. F_E is independent of membership of category C , but provides an alternative explanation of the most diagnostic of the original features, F_{1-3} .

For example, we might tell people about an animal that has four legs, a tail, and is very smelly. In this baseline condition, we would ask people to rate the perceived likelihood that this creature is a skunk. In the experimental condition, we tell people about an animal that has the same set of features—four legs, a tail, and has an unpleasant odor—but *in addition* tell them that the animal has just been wading through a sewer. Although wading through a sewer is a behavior not associated specifically with skunks (other animals are equally likely to be found rummaging around sewers) and therefore should not directly affect people's confidence that the animal is a skunk, the behavior is nonetheless causally associated with one of the features. Wading through sewers can cause an animal to gain a foul stench. Thus in the experimental condition, there is a cause aside from category membership that explains a highly diagnostic feature. Our model therefore predicts that participants in the experimental condition will reason that the animal is less likely to be a skunk.

Of course, there is another competing explanation to this pattern of results. It could simply be that adding additional information to the scenario gives people more information to assimilate and hence makes them inherently less confident in their predictions. In order to address this, we included a control condition in which an extra piece of information was added that was unrelated to category membership, and did *not* undermine diagnostic features. In the control condition, the additional feature, F_C , is independent of membership of category C , and also causally unrelated to the original features, F_{1-3} . If, as our model predicts, category irrelevant information only affects categorization confidence when it is causally related to one of the baseline features, then we would expect the baseline and control ratings to be indistinguishable from one another. The design of the stimuli in the three conditions is illustrated in Figure 6.5.

Many studies have shown that there are differences between natural kind and artifact categories (e.g., Barton & Komatsu, 1989; Bloom, 1998; Lingle, Alton, & Medin, 1984; Rips, 1989). Medin, Lynch, and Solomon (2000) have suggested that the types of features used in categorization of artifacts differ from the types of features used to categorize natural kinds. In particular, features that influence function are more important for artifact categories, while features about internal structure are more important for natural kind categories. The fact that different types of features are causal for natural kind and artifact categories (Keil, 1995), may explain the differences in the central features of natural kind and artifact categories.

However, in our model, we presume that the category itself is a central cause, in that category membership causes *all* other features. Thus, our predicted pattern of results should hold regardless of whether a category is a natural kind or an artifact.

Some researchers have also argued that social categories may be different from object categories (Lingle et al., 1984; Medin et al., 2000). Social categories are perceived as more flexible, and categorization is often strongly influenced by individuating information (Fiske, Lin, & Neuberg, 1999; Kunda & Thagard, 1996). However, according to our model, non-diagnostic individuating information should not reduce the likelihood of discounting, and thus the predicted pattern of our model also should hold regardless of whether a category is a social category or an object-based category.

As such, in the experiments below we tested three types of categories: natural kind, artifact, and social categories.

4.2. Predictions of the Model

The sigmoidal model predicts patterns of discounting and augmenting across the full range of parameter settings, without requiring that parameters be tuned to any precise values. These predictions can be shown to hold analytically, but we also verified them through extensive simulations of the model. We varied the prior probabilities of the causes from 0.000001 to 0.99999, and varied the causal strengths such that the probability of the effect given each cause varied from 0.000001 to 0.99999. The baseline probability of the effect (with no causes present) was held constant at 0.1, although we have found qualitatively similar results at other settings of this parameter. In all cases, regardless of the prior probabilities or causal strength parameters, these simulations confirmed that the model produces discounting for generative alternative causes and augmenting for preventive alternative causes.

4.3. Do Category-Irrelevant Contextual Features Affect Categorization

We created a set of features to describe natural kind categories (ostrich, bear, and skunk), mechanical categories (golf cart, refrigerator, and stamp) and social categories (oil baron, firefighter, and cheerleader). For example, a skunk was described as having four legs, a tail, and being smelly. An additional feature was then created that was unrelated to the category but could explain the most diagnostic of the original features; rummaging in the sewers is independent of “skunkness”, but could explain the creature’s

smell. Finally, a decoy feature completely unrelated to either the category or any of the other features was created to serve as a control (“has just eaten” in the skunk example). Separate norming studies were run to ensure that discounting and nondiagnostic features were indeed judged to be independent of the target categories. Table 6.1 provides a full list of features used in this experiment.

Sixty-two undergraduate participants were provided with three categories and a list of features for each category. One third of the participants were provided with the three main features to serve as a baseline for category judgments. Another third of the participants were in the experimental condition, which included the discounting feature in addition to the main features. The final third were in the control condition and were given the three main features and the decoy feature. In all conditions, participants were asked to rate how likely each exemplar was to be a member of the category using a 0 (definitely is not a member) to 10 (definitely is a member) scale. The model predicts that participants will provide lower evaluations of

Table 6.1 List of Categories, Features, Key Features, Category-Irrelevant Features Causally Related to Diagnostic Features, and Category-Irrelevant Features Unrelated to Diagnostic Features for the First Study

Category	Baseline Features	Target Feature	Experimental Feature	Control Feature
Ostrich	Feathers, lay eggs	Cannot fly	Broken wing	Standing in the sun
Bear	Fur, teeth	Attacks anything near it that moves	Rabies	Near a tree
Skunk	Four legs, a tail	Smelly	Rummaging in sewers	Has just eaten
Golf Cart	Four wheels, Carries people	Top speed of 30mph	Badly needs repairs	Badly needs a wash
Refrigerator	Shelves, things inside	Is cold inside	Located in Alaska	Located in Hawaii
Stamp	Small, paper	Sticky	Orange juice spilled on it	Is in a house
Firefighter	Physically fit, Brave	Wears a uniform	Plays amateur football	Has a daughter
Oil Baron	Southern drawl, owns large house	Extremely rich	Lottery winner	Eats chocolate
Cheerleader	Blonde, wears short skirt	Hyperactive	Drinks espresso	Lives in Phoenix

category membership for the experimental condition than for the baseline and control conditions.

For the natural kind categories, participants rated the animals as better category members for the baseline condition ($x = 6.20$) and the control condition ($x = 6.35$) than for the experimental condition ($x = 4.54$). An ANOVA was run on the data with a planned contrast of $-1, -1, 2$, and showed that this difference was significant ($t(183) = -4.7, p < 0.001$).

Similar trends were found for artifact categories. On average, participants rated the artifacts as better category members for the baseline condition ($x = 7.23$) and the control condition ($x = 7.22$) than for the experimental condition ($x = 4.91; t(183) = -6.2, p < 0.001$).

For social categories, participants also rated the exemplars as better category members for the baseline condition ($x = 6.19$) and the control condition ($x = 5.59$) than for the experimental condition ($x = 4.96; t(415) = -3.6, p < 0.001$). It is worth noting that unlike for the natural kind and artifact categories, for social categories the control condition lays in between the baseline and experimental conditions. While this difference was not significant, it nonetheless conforms to the notion that nondiagnostic individuating information has a larger influence on social categories than object categories (Fiske et al., 1999; Kunda & Thagard, 1996; Medin et al., 2000).

This pattern of data conforms to what one would expect from the discounting hypothesis. In the experimental condition, participants judged the features to be significantly less indicative of category membership than in the baseline condition. Further, because the control and baseline condition did not reliably differ, it seems unlikely that this result is due to the addition of an irrelevant cue lowering participants' confidence. Category judgments only decreased when an alternative cause for the diagnostic feature was provided.

4.4. Can Category-Irrelevant Contextual Features be Preventative as well as Generative?

Although the study described above focused on the influence of category-irrelevant *generative* causes for a diagnostic feature, our model makes predictions about category-irrelevant *preventative* causes as well. In particular, if a diagnostic cue exists in the presence of a category-irrelevant cause that lessens the likelihood of that cue, our model predicts that judgments of category membership should increase. That is, if we learn that an animal has been sprayed with a pleasant perfume, and *still* smells bad, it should increase our confidence that the animal is a skunk. This prediction is in line with Kelley's (1972) notion of augmenting; given the fact that a pleasant odor has been externally applied to our malodorous

animal and the animal still smells, another cause of the stench (such as being a skunk) must exist. Thus when a diagnostic cue exists in the presence of a preventative cause, our confidence in category membership should increase.

To test this, 45 additional undergraduates took part in a second study. As in the first study, participants were given a list of three features, F_1 , F_2 , F_3 , which were diagnostic of a category C. However, in the second study the feature list was preceded by a vignette. These vignettes were interesting stories which contained many neutral category irrelevant details. At the end of the stories, a protagonist encounters some object, and a list of features describing the object. As in the first study, the participants were asked to use a 10-point scale to evaluate category membership.

For example, the following vignette preceded the categorization judgment for the social category “cheerleader”:

Lisa is house-sitting for friends who are out of town. She has lost their mail, forgotten to feed the dog, and has just swallowed sleeping pills from their medicine cabinet. Lisa

- is hyperactive
- wears a short skirt
- is blonde

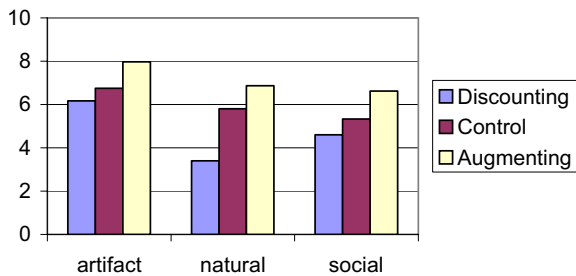
In this vignette, the key feature is “swallowed sleeping pills” which is a preventative of hyperactivity. In the discounting condition, Lisa swallowed caffeine pills, while in the baseline condition there was no mention of pills at all. A list of categories, features, and causally relevant details can be found in [Table 6.2](#).

For all three category types, the pattern of results was the same; participants rated the exemplars in the baseline condition (natural kind = 5.80, artifact = 6.75, social = 6.33) as poorer category members than in the augmenting condition (natural kind = 6.87, artifact = 7.97, social = 6.61) and as better category members than in the discounting condition (natural kind = 3.40, artifact = 6.17, social = 4.60). These differences were reliable for all three category types (natural kind: $F(42,2) = 7.6, p < 0.05$; artifact: $F(42,2) = 2.6, p < 0.05$; social: $F(42,2) = 4.2, p < 0.05$) and for the data set as a whole ($F(42,2) = 7.7, p < 0.05$). The results are summarized graphically in [Figure 6.6](#).

The patterns in the data conformed exactly to our model’s prediction. When a category-irrelevant *generative* cause for a diagnostic feature is present, people’s judgments of category membership decrease, but when a category-irrelevant *preventative* cause for a diagnostic feature is present, people’s judgments of category membership increase.

Table 6.2 List of Categories, Features, and Generative and Preventative Causes Used in the Second Study. Key Features Are Listed First and in Italics

Category	Features	Augmenting Detail	Discounting Detail
Skunk	<i>Smells bad</i> Four legs Tail	Sprayed with pleasant smelling chemical agent	Sprayed with unpleasant smelling chemical agent
Venus fly trap	<i>Moving</i> Green Six inches tall	Still day	Windy day
Refrigerator	<i>Cold</i> Has shelves Has things inside	In a scorching desert	In an arctic wasteland
Sports car	<i>Going 85mph</i> Two-door Painted white	Going uphill	Going downhill
Cheerleader	<i>Hyperactive</i> Blonde Wearing short skirt	Has taken sleeping pills	Has taken caffeine pills
Math genius	<i>100% on midterm</i> Wears glasses Has brown hair	Did not study for the midterm	Studied very hard for the midterm

**Figure 6.6** Results of study 2. Category-irrelevant generative causes for diagnostic cues led to discounting, while preventative cues led to augmenting. For a color version of this figure, the reader is referred to the online version of this book.

5. PLACING THE PRESENT ACCOUNT IN THE LITERATURE

5.1. Other Accounts of Discounting

While we have argued for a Bayesian account of categorization as causal explanation, there are other mathematical models in the causal reasoning literature that predict discounting and augmenting. However, these models

currently have trouble incorporating our findings with those from the categorization literature.

For example, constraint satisfaction models (Holyoak & Thagard, 1989; Thagard, 1989, 1992) have been applied successfully to the problem of causal attribution and more specifically to discounting (Miller & Read, 1991; Read, 1987; Read & Marcus-Newhall, 1993). These models account for discounting in two ways. The first is through the use of inhibitory links between potential causes (Read & Marcus-Newhall, 1993). However, this does not describe our data, in which alternate causes are independent or positively related. The second approach to discounting is normalizing the total amount of activation in the network. However, this would predict that increasing the number of features associated with a category would actually *decrease* category confidence. As more nodes representing features are activated (e.g., “smells”, “has tail”, “is a mammal”) the more “skunk” has to compete for scarce resources, and the lower its activation will be. This prediction goes against most of the findings from the feature-based categorization literature (e.g., Nosofsky, 1986).

Another computational account of causal discounting can be found in feedforward connectionist nets (McClelland & Rumelhart, 1988; Van Overwalle, 1998). Van Overwalle (1998) and Van Overwalle and Van Rooy (2001) argue that feedforward networks account for discounting during the learning stage. When multiple causes are active in the presence of an event, they have to compete for the available adjustments of link weights. However, when there is only one cause active, there is no such competition, and the nonpresent cause is given a lower weight relative to the present cause. Although some researchers have argued that this inhibitory effect in causal learning is equivalent to discounting (Baker, Mercier, Vallee-Tourangeau, Frank, & Pan, 1993; Shanks, 1991), there does seem to be a distinction (Read & Montoya, 1999): if a person has already learned that skunks and sewage are legitimate causes for an unpleasant odor, then the model cannot explain how the presence of sewage might decrease that person’s confidence in the presence of a skunk as was found in our studies.

Another approach to discounting is Medcof’s (1990) Probability Expectancy Attribution Theory (PEAT), which argues that the strength of causal attribution is divided by the total number of causes present. However, this ratio rule cannot incorporate the relative plausibility of alternate causes into the discounting process. The number of events present is computed without regard for how reasonable it is to ascribe causation to those events. Further, Einhorn and Hogarth (1986) have noted

that ratio models cannot account for low causal ascription in the absence of alternate causes.

In sum, while there are a number of models that can predict discounting (for a review, see [Khemlani & Oppenheimer, 2011](#)), they are limited in their ability to apply those predictions to categorization, unlike our Bayesian network models which subsume several important previous approaches and thereby capture a broad range of categorization phenomena. Undoubtedly, other approaches could be extended to do this and doing so might yield interesting predictions. However, we believe that our approach is a particularly elegant way of treating these issues because it so naturally extends previous causal and noncausal accounts of categorization as well as capturing the new discounting and augmenting effects we have presented here.

5.2. Relating the Bayesian Network Model to Other Models of Causally Based Categorization

Until now, we have emphasized the points of contrast between our account and other models of causally based categorization in order to highlight its distinctive contributions. Ultimately, we see our model more as a complement to existing models of categorization as causal reasoning, rather than as a competitor. While most other theories look at how features within a concept are causally related, ours examines how categorization interacts with context and background knowledge more holistically. Taken in conjunction, these two approaches to causally based categorization provide a great deal of explanatory reach.

For example, [Ahn's \(1998\)](#) theory of feature centrality suggests that some features are more central to some categories than others. The extent of discounting will likely vary based on the diagnosticity of a given feature for category membership. That is, category-irrelevant causes for features will have a greater effect on eventual categorization decisions if they explain features that in turn cause other features.

Similarly, [Rehder's \(2003\)](#) causal model theory examines causal relations within a concept. The specific Bayesian network model we tested in our experiments here assumes the features are all conditionally independent given the category membership. While this simplifying assumption makes the model more readily testable, it is clearly too simple to capture many kinds of natural categories. Combining a representation of category-internal causal structure with external causal information should lead to a richer and more accurate Bayesian network account of categorization.

The original motivation for our framework came from [Murphy and Medin's \(1985\)](#) example of categorizing as “drunk” a person who jumps fully clothed into a swimming pool at a party. This judgment clearly draws on knowledge external to the category *drunk*. To account for this judgment using prior theories of categorization, one would need to stipulate that within their concepts of *drunk*, people store knowledge that being drunk can cause a man to jump into a pool fully clothed (see model depicted in [Figure 6.7](#)), which most researchers consider to be unreasonably stretching the notion of conceptual knowledge. A further wrinkle is that if the man saw someone drowning before jumping in, it is unlikely that one would infer that he is drunk. This inference is not possible without including in the model other possible causes of jumping into a pool fully clothed.

To illustrate how an extended version of our framework could account for this intuition, [Figure 6.8](#) depicts a causal network that embeds *drunk* within a model that includes external knowledge. This model is simplified to make the computations comprehensible (e.g., we do not model the fact that wanting to swim should cause a person to not be fully clothed), yet it effectively explains several important intuitions. According to this model, the concept of drunk need only contain knowledge that being *at a party* can cause one to become *drunk*, and being *drunk* can cause one to be *uninhibited*. This knowledge can then be combined with knowledge of causes for jumping into a pool. In this model, *wanting to swim* and *seeing someone drowning* can both cause one to jump into a pool, while being *fully clothed* inhibits one from jumping into a pool. This model, properly parameterized, predicts the inference that a man jumping into a pool fully clothed at a party is drunk. Being at a party increases the prior probability that a man would be drunk, and being drunk would make the man uninhibited, thus weakening the mechanism by which *fully clothed* inhibits *jumps in pool*. Of course, constructing this model requires specific knowledge that the mechanism by which *fully clothed* inhibits *jumps in pool* is psychological, and depends on the man having the ability to inhibit his impulses, which being drunk impairs.

This model can also account for intuitions about this situation beyond the one that [Murphy and Medin \(1985\)](#) focus on. If all we know is that a man jumped into a pool, we might infer that he wanted to swim. If it is further known that the man saw someone drowning, the model would predict discounting of *wants to swim*. The model would also predict augmenting of *sees someone drowning* if it is known that a man jumped into a pool fully clothed but was not drunk, and would predict discounting of *drunk* if it were known that the man saw someone drowning.

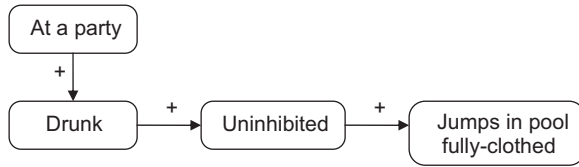


Figure 6.7 Example causal model of the concept “drunk”, restricted only to features of the concept drunk, designed to account for the intuition that a fully clothed man jumping into a pool at a party is drunk. Using prior causal model theories of categorization, one could only account for this intuition by including in one’s concept of drunk the knowledge that drunken people are likely to jump into a pool fully clothed. This was argued to be implausible by [Murphy and Medin \(1985\)](#).

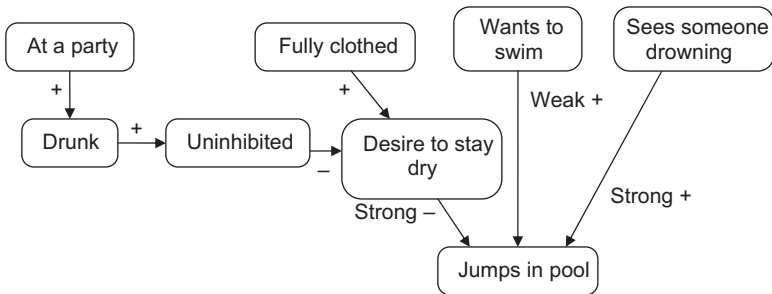


Figure 6.8 Example causal model which embeds the concept “drunk” within a larger set of causes, to account for the intuition that a fully clothed man jumping into a pool at a party is drunk. This model only includes causal links that are plausibly consistent with common knowledge, and explains the intuition without requiring the concept of drunk to contain knowledge about jumping into a pool fully clothed.

5.3. Caveats

There is much about the cognitive basis of categorization that our model does not attempt to address. We are not arguing that Bayesian networks capture how abstract knowledge about categories or causal relations is represented in long-term memory, or how categories are learned in the first place. We are not modeling intuitive theories or the structure of concepts—deep and important but very hard problems. Rather we have presented a model of categorization.

Moreover, we have not attempted to account for all of the many phenomena that have been demonstrated in the categorization literature (for a review of knowledge effects in categorization, see [Murphy, 2002](#)). For example, [Barsalou’s \(1991\)](#) work on ad hoc categories and [Wisniewski’s \(1995\)](#) work on how prior knowledge helps constrain what features are used in categorization are both highly relevant, but beyond the scope of

the present discussion. Ultimately, Bayesian networks can be constructed in working memory based on a variety of types of knowledge from long-term memory, and can include features caused by the category as well as category-irrelevant events. These networks can be used to support online judgments about category membership in the presence of several causal cues, but the processes by which an individual builds these networks in real time, drawing on just the right background knowledge, remain to be explored in future work.



6. CONCLUSION

The world is a complex place, and categories (and their features) are embedded within a web of causal connections. In this paper, we have attempted to formally model how people might understand interactions between category features and the broader external context. Based upon the predictions of this model, we have shown that information typically thought of as category-irrelevant can nonetheless influence people's causal understanding of more category-relevant features. As researchers continue to explore how causal knowledge impacts reasoners' understanding of concepts and categories, it will be important to further examine how people incorporate their rich knowledge about the world in which those concepts and categories reside.

ACKNOWLEDGMENTS

We thank Dan Yarlett for very generous support, feedback, and assistance; without his aid this work would not have been possible. We would also like to thank David Collister, Nicolas Davidenko, Benoît Monin, Joel Allan, Sara Etchison, Matthew Weber, and Mariam Hambarchyan for their very helpful comments and suggestions. DMO and TRK were supported by National Science Foundation Graduate Research Fellowships. DMO was supported by NSF 1128786. JBT was supported by the Paul E. Newton Career Development Chair.

REFERENCES

- Ahn, W. (1998). Why are different features central for natural kinds and artifacts?: the role of causal status in determining feature centrality. *Cognition*, *69*, 135–178.
- Ahn, W. (1999). Effect of causal structure on category construction. *Memory and Cognition*, *27*, 1008–1023.
- Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, *41*, 361–416.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

- Atran, S. (1987). Origin of the species and genus concepts: an anthropological perspective. *Journal of the History of Biology*, 20, 195–279.
- Baker, A. G., Mercier, P., Vallee-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective associations and causality judgments: presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 414–432.
- Barselou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 27, pp. 1–64). New York: Academic Press.
- Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. *Journal of Psycholinguistics Research*, 18, 433–447.
- Bloom, P. (1998). Theories of artifact categorization. *Cognition*, 66, 87–93.
- Danks, D. (2007). Theory unification and graphical models in human categorization. In A. Gopnik, & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford: Oxford University Press.
- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, England: Cambridge University Press.
- Einhorn, H., & Hogarth, R. (1986). Judging probable cause. *Psychological Bulletin*, 99, 3–19.
- Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken, & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 231–254). New York: Guilford.
- Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: a framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 234–257.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford: Oxford University Press.
- Gelman, S. A., & Welman, H. M. (1991). Insides and essences: early understandings of the nonobvious. *Cognition*, 28, 213–244.
- Gluck, M. A., & Bower, G. H. (1988a). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General*, 117, 227–247.
- Gluck, M. A., & Bower, G. H. (1988b). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, 27, 166–195.
- Gopnik, A., & Meltzoff, B. (1997). *Words, thoughts, and theories*. Cambridge, MA: Bradford, MIT Press.
- Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295–355.
- Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: a multidisciplinary debate*. Oxford: Oxford University Press.
- Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: perceiving the causes of behavior* (pp. 151–174). Morristown, NJ: General Learning Press.
- Khemlani, S., & Oppenheimer, D. (2011). When one model casts doubt on another: a levels-of-analysis approach to causal discounting. *Psychological Bulletin*, 137, 195–210.
- Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: a parallel-constraint-satisfaction theory. *Psychological Review*, 103, 284–308.
- Lingle, J. H., Alton, M. W., & Medin, D. L. (1984). Of cabbages and kings: assessing the extendibility of natural object concept. Models to social things. In R. S. Wyer, & T. K. Srull (Eds.), *Handbook of social cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Locke, J. (1894/1975). *An essay concerning human understanding*. Oxford: Oxford University Press.

- Lord, P., & Park, N. (2000). *Chicken run*. : Dreamworks, SKG.
- McClelland, J. M., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: a handbook of models, programs and exercises*. Cambridge, MA: Bradford Books.
- Medcof, J. W. (1990). PEAT: an integrative model of attribution processes. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 111–209). San Diego, CA: Academic Press.
- Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology*, 51, 121–147.
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge, MA: Cambridge University Press.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.
- Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, 20, 158–190.
- Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 241–253.
- Miller, L. C., & Read, S. J. (1991). On coherence of mental models in persons and relationships: a knowledge structure approach. In F. Fincham, & G. J. O. Fletcher (Eds.), *Cognition in close relationships*. Hillsdale, NJ: Erlbaum.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge: MIT Press.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 87–108.
- Nosofsky, R. M. (1991). Relation between the rational model and the context model of categorization. *Psychological Science*, 2, 416–421.
- Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (pp. 149–167). Hillsdale, NJ: Erlbaum.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Palo Alto: Morgan Kaufmann Publishers.
- Pearl, J. (2000). *Causality—models, reasoning, and inference*. Cambridge: Cambridge University Press.
- Putnam, H. (1977). Is semantics possible? In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds*. Ithaca, NY: Cornell University Press.
- Read, S. J. (1987). Constructing causal scenarios: a knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology*, 52, 288–302.
- Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: a parallel distributed processing account. *Journal of Personality and Social Psychology*, 65, 429–447.
- Read, S. J., & Montoya, J. A. (1999). An autoassociative model of causal reasoning and causal learning: reply to Van Overwalle's (1988) critique of Read and Marcus Newhall (1993). *Journal of Personality and Social Psychology*, 76, 728–742.
- Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382–407.
- Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science*, 27, 709–748.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge: Cambridge University Press.
- Rossee, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, 46(2), 178–210.

- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. *Proceedings of the 28th annual conference of the cognitive science society*.
- Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 7, 433–443.
- Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, 12, 435–467.
- Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.
- Van Overwalle, F. (1998). Causal explanations as constraint satisfaction: a critique and a feed-forward connectionist alternative. *Journal of Personality and Social Psychology*, 74, 312–328.
- Van Overwalle, F., & Van Rooy, D. (2001). How one cause discounts or augments another: a connectionist account of causal competition. *Personality and Social Psychology Bulletin*, 27, 1613–1626.
- Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121, 222–236.
- Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General*, 124, 181–206.
- Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *JEP: LMC*, 21, 449–468.

This page intentionally left blank



Individual Differences in Intelligence and Working Memory: A Review of Latent Variable Models

Andrew R. A. Conway^{*,1}, Kristof Kovacs^{}**

^{*}Department of Psychology, Princeton University, Princeton, NJ, USA

^{**}Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

¹Corresponding author: E-mail: aconway@princeton.edu

Contents

1. Introduction	234
1.1. Overview	234
1.2. Two Disciplines of Psychology	235
1.3. Unification: Benefits and Challenges	237
1.4. Interim Summary	240
2. Models of Intelligence	240
2.1. Factor Analysis	240
2.2. A Single General Factor	241
2.3. Several Group Factors	242
2.4. Higher-order and Bi-factor Models	244
2.5. The Content of Factors	246
3. Interpretation of the General Factor	247
3.1. A Unitary Source	247
3.2. Evidence Against a Unitary Source	249
3.3. Multiple Sources	250
4. Models of Working Memory Capacity	252
4.1. History Repeating	252
4.2. The General Factor, Again	255
5. Toward a Unified Model	256
5.1. Intelligence and Working Memory Capacity	256
5.2. Components of Working Memory	257
5.3. Components of Intelligence	258
5.4. Converging Evidence	259
6. Conclusion	262
Appendix	263

Abstract

The purpose of the current review is to examine individual differences in intelligence and working memory capacity. The emphasis is on latent variable models and theoretical frameworks that connect interindividual differences in behavior with intraindividual psychological processes. Our review suggests that intelligence and working memory capacity are strongly correlated and that the shared variance is primarily due to the fluid reasoning component of intelligence and mechanisms of cognitive control in working memory. We conclude that research on intelligence and working memory is a rare successful example of the unification of experimental and differential psychology. Finally, we argue that general ability models of intelligence that posit a unitary source of variance are not consistent with contemporary research and should be fairly rejected.



1. INTRODUCTION

1.1. Overview

Intelligence and working memory are among the most extensively studied constructs in Psychology. Both are related to higher-level cognition, both involve processes associated with cognitive control, and both have been linked to critical neural mechanisms in pre-frontal cortex. It should therefore come as no surprise that the connection between these constructs has become a major focus of research interest in cognitive psychology and neuroscience.

Search of PsycInfo, an online database, carried out on June 30, 2012 showed that the number of journal articles and dissertations that included both the word ‘Intelligence’ and the phrase “working memory” in the abstract or title has consistently increased over the last 30 years. In successive 5-year periods between 1981 and 2010, the numbers were 7, 15, 37, 87, 218, and 426! Not only did the first decade of the third millennium see a huge increase in joint research on intelligence and working memory, the years between 2006 and 2010 produced more papers and dissertations on this topic than the previous 25 years combined.

Given this recent proliferation of research, a single manuscript cannot possibly review all contemporary work. Therefore, our goal here is to review latent variable models of intelligence and working memory, and in so doing, to address two of Psychology’s most vexing questions: (1) is it possible, or even desirable, to bridge experimental psychology and differential psychology? (2) What is the theoretical status of the general factor in models of cognitive ability?

1.2. Two Disciplines of Psychology

At the outset, it is essential to highlight a fundamental difference between intelligence and working memory. After a few unsuccessful early attempts to channel research on individual differences in mental abilities into mainstream experimental psychology (e.g. Cattell, 1890), and following the invention of the modern IQ-test at the turn of the 20th century, the field of intelligence was essentially agnostic with respect to intraindividual psychological processes, and for most of the twentieth century focused solely on the measurement and explanation of *interindividual differences* in cognitive abilities (Mackintosh, 2011). Working memory, on the other hand, is a construct developed by cognitive psychologists to refer to the *intraindividual processes* that enable one to hold goal-relevant information in mind, often in the face of concurrent processing and/or distraction (Baddeley, 1992; Baddeley & Hitch, 1974).

Thus, joint research on intelligence and working memory not only connects two fields of study, but is also an ideal candidate to answer Cronbach's (1957) call for integrating the two disciplines of scientific psychology, that is, the one studying interindividual differences ("differential psychology") and the one studying intraindividual processes ("experimental psychology"). Cronbach's call for a unified psychology has been echoed on several occasions (e.g. Underwood, 1975) as well as for different target audiences: intelligence researchers have been encouraged to apply experimental methods (Eysenck, 1995), and experimental psychologists have been advised to favor nomothetic theories that also explain individual differences (Cohen, 1994). Yet, unified models are still extremely rare in psychology because bridging the two disciplines presents unique challenges, which we will discuss in more detail below. As well, in many realms of psychology it is possible to maintain a successful program of research by working in one discipline while ignoring the other.

One area where this is not possible is developmental psychology. Theories of development can't dismiss interindividual differences in behavior as noise. For example, despite having enormous impact, Piaget's theory of cognitive development has been widely criticized because it failed to explain why some children advanced to different stages at different rates. In the 1970s, developmental psychologists began to address this limitation by incorporating ideas from information processing models of cognition (Case, Kurland, & Goldberg, 1982; Pascual-Leone, 1970). This work, in turn, had an enormous impact on the field of working memory and set the stage for a unified account of intelligence and working memory (Case & Daneman, 1981; Daneman & Carpenter, 1980).

A pivotal moment was the introduction of so-called *complex span tasks*, that is, tasks which require the simultaneous storage and processing of information. The tasks are unique because they are developed from a theoretical perspective—Case’s theory of development and Baddeley’s theory of working memory—and they account for important individual differences in cognitive performance. One of the first such tasks, reading span, in which subjects have to read sentences and remember the last word of each sentence, was developed by Daneman and Carpenter (1980), who showed that variation in performance on this particular measure of working memory capacity predicts performance on reading comprehension tests. Over the years, many versions of complex span tasks have been developed (for a review of working memory span tasks, see Conway et al., 2005). The point here is that the construction of complex span tasks is a theory-driven enterprise. In contrast, the construction of intelligence tests is largely a data-driven enterprise, that is, tests are constructed to predict various real-life outcomes. Connecting complex span tasks to intelligence tests therefore connects theory to data in a novel way.

Amid the siren songs calling for the unification of differential and experimental psychology, a few researchers have expressed skepticism. Jensen (2000) argued that differential and experimental psychology are justifiably separated, because in order to systemize individual differences, one need not understand universal laws of cognition. Contrary to Jensen, who appears to be in favor of the separation of the two disciplines, Borsboom, Kievit, Cervone, and Hood (2009), albeit not having much faith in convergence either, do allow special conditions under which a common framework is tenable: “the idea that correlational and experimental research can ‘converge’, in the sense that they render support for the *same* hypothesis (...) only makes sense in a limited set of situations—namely those in which the interindividual differences found in correlational research are *exclusively* the result of the intraindividual processes studied in the corresponding experimental research” (p. 73). Moreover, “(t)o have a real connection between the fields under consideration here, one should be able to infer what an interindividual differences structure should (sic!) like from a theory of intraindividual processes—more specifically, one should be able to place refutable restrictions on the interindividual model structure” (p. 93).

Borsboom et al. come to the conclusion that these conditions have not yet been appropriately met, and quite likely never will be met. While agreeing with the conditions they propose, we are substantially more optimistic about future possibilities of integrating the differential and experimental

disciplines, in particular, of integrating models of individual differences in intelligence with intraindividual theories of working memory, primarily through the interindividual construct “working memory capacity.” It further sparks our optimism that Cohen (1994) cites Baddeley and Hitch’s (1974) model of working memory as “an excellent example of how experimental data and individual differences can be treated within the same theoretical framework” (p. 10), and even Borsboom et al’s highly pessimistic evaluation admits that “it would seem that the experimental psychologists (sic!) ‘working memory’ and the differential psychologists (sic!) ‘working memory’ are related, and how they may be is an important issue” (p. 94).

Besides skepticism about a common framework of experimental and correlational psychology, it appears that Borsboom et al. (2009) and Jensen (2000) also share a fondness of aliens and car analogies. They both apply a similar thought experiment to show that if extraterrestrial creatures were to study cars, they would most likely develop separate lines of research for (a) how individual cars work, and (b) what makes them differ in performance. While these thought experiments seem convincing, there would be practical problems that our alien would have to face. Let us assume that she is interested in the speed of cars and she finds that faster cars have, in general, more gears, larger engines, more valves per cylinder, and fancier hubcaps. Which of these *cause* some cars to go faster than others?

A radical improvement in the state of affairs in differential psychology is that, thanks to the advancement of confirmatory factor analytical techniques, one can now test elaborate causal models about individual differences—something that, at the time Cronbach’s paper appeared, was considered to be the privilege of experimental psychologists working from nomothetic theories. In our example, one could rule out the possibility that hubcaps causally influence speed solely by applying statistical control. Nevertheless, converging lines of evidence can always further corroborate this conclusion, such as effects of experimental manipulations (do fitting different hubcaps significantly influence performance?) or other natural causes (does performance decrease as a result of losing a hubcap or does it remain intact?).

1.3. Unification: Benefits and Challenges

Even though the conditions for research on intraindividual processes and individual differences to converge under the same theoretical framework are hard to meet, when they in fact can be met, the advantages are enormous. A theory with more converging lines of evidence is undoubtedly better supported than one with less. As well, a theory of individual differences

which is not embedded into a theory of within-individual processes, and which therefore cannot use explanatory constructs discovered by individual-level research, is scientifically inferior to a theory which is embedded in such a theoretical framework. In the car example, an account of why some cars are faster than others that is based on the explanation of how engines work is necessarily more elaborate, plausible, and parsimonious than one that only explores variation, attempting to explain speed differences without an understanding of what makes speed possible in the first place.

The advantages of differential/correlational psychology as an individual discipline are therefore contingent upon the development of theories that provide psychological explanations and predictions, rather than mere structural accounts of covariance. As we shall see, there are occasions when different models provide equally good fit to interindividual data, but these models are compatible with different accounts of intraindividual processes. Without taking into account these processes, and relying only on individual differences data, such models are technically isomorphic.

Furthermore, the interpretation of latent variables is a very difficult enterprise, which without corroborating evidence about intraindividual processes is restricted to the analysis of test content and regression parameters of the model. Many cognitive scientists are skeptical about the objectivity of this process, the most pronounced of whom is probably Stevan Harnad, who claims that: “there is a huge hermeneutic component to psychometric analysis. The empirical part is the calculation of the correlations in the extraction of factors; the hermeneutic part is in interpreting the factors, figuring out what on earth they may mean” (Bock, 2000, p. 48).

In spite of several warnings (e.g. Borsboom, Mellenbergh, & Van Heerden, 2003; Jensen, 1998) the vast majority of differential psychologists are apparently unable to resist the temptation to carelessly link latent variables to intraindividual processes. The foremost example is the interpretation of psychometric *g*, the general factor obtained from diverse batteries of mental tests, as “general intelligence”, a domain-general intraindividual mechanism that enables *any given individual* to solve items in IQ-tests. In fact, the very practice of labeling tests with completely different content as “IQ-tests” is a case of implicitly assuming that all tests tap the same domain-general mechanism.

For these reasons we believe that the convergence of both interindividual and intraindividual lines of working memory research as well as the study of variation in intelligence is a meaningful and promising enterprise. However, we admit that this has to be done with exceptional rigor. That is,

when evaluating converging models of intelligence and working memory, we will take very seriously the conditions under which such convergence makes sense.

First, it is important to be aware that evidence about individual differences and intraindividual processes are mutually underdetermined with regard to theories of intraindividual processes and the structure of individual differences, respectively. This means that, on the one hand “even the most impressive fit of a between-subjects model to interindividual differences data does not have implications for the structure of psychological attributes or processes that operate at the level of the individual” (Borsboom et al., 2009, p. 76.). In the car example, studying how variation in some components causes variation in the speed of the car will never reveal how individual cars obtain their speed. On the other hand, even though a theory of intraindividual processes is capable of predicting the structure of individual differences, “the set of situations in which laws concerning universal processes yield any predictions about the structure of interindividual differences is highly limited” (Borsboom et al., 2009, p. 73). Once again, in the car analogy: (a) some components either do not manifest variation, or their variation does not correlate with overall performance (such as that of hubcaps) and (b) differentiable components can still correlate and thus load on the same latent variable, even to such an extent that makes them unitary from an individual differences perspective (such as power and capacity).

Second, from the mutual underdetermination of individual differences and within-individual data to the other type of theory follows the necessity of the distinction between intraindividual and interindividual concepts. Such a distinction, however, is rarely made, which results in conceptual confusion when the two disciplines actually communicate. Differential psychologists’ concept of “general intelligence”, even if it is simply meant as “the general factor of intelligence”, and never in fact refers to an intraindividual mechanism, is still naturally interpreted by cognitive psychologists and cognitive scientists as a domain-general, within-individual cognitive mechanism: something like, for instance, Newell and Simon’s (1972) “General Problem Solver.” This interpretation, in turn, sharply opposes domain-specificity and modularity, whereas the general factor of intelligence clearly does not, since it only says that variation appears to be domain-general at the group level. At the individual level, it can be completely domain-specific (for a theory that actually incorporates both modularity and a general factor of mental abilities, see Anderson, 1992). With respect to the topic of this paper, it is crucial to distinguish working memory as an intraindividual and as an

interindividual construct. To do so, we will refer to “working memory” and “working memory capacity”, respectively.

Third, it is imperative to keep in mind Borsboom et al.’s call for *refutable* predictions from intraindividual theories to the structure of individual differences. This suggestion is especially important in the current state of research on intelligence and working memory: since it has been firmly established that they are strongly related constructs, research on the nature of the relationship could enter a more mature stage by specifying which components are responsible for the correlation and, more importantly, which are not. “Every ‘good’ scientific theory is a prohibition: it forbids certain things to happen. The more a theory forbids, the better it is” (Popper, 1963, p. 36). A strong theoretical account of intelligence and working memory will therefore predict which outcomes will be correlated *and* it will predict which outcomes are *not* correlated.

1.4. Interim Summary

To sum up this rather lengthy Introduction: the purpose of the current chapter is to closely examine latent variable models of intelligence and working memory capacity and then determine whether a unified model is possible. While doing so, we will prefer individual-differences models that are predicted by accounts of intraindividual cognitive processes, and we will prefer refutable, differential predictions to cases of simple convergence.

First, we discuss accounts of the structure of individual differences in intelligence. Second, we try to do justice to explanations of such structure in general, and in particular as to whether the general factor of intelligence is the result of a unitary process. Third, we provide a similar outline of working memory capacity, focusing on whether a domain-general account of variance is tenable. This is followed by an evaluation of whether such domain-general variance in working memory capacity is the result of a unitary mechanism or a number of interacting components. Finally, we discuss how models and interpretations of intelligence and working memory might converge to a unified account.



2. MODELS OF INTELLIGENCE

2.1. Factor Analysis

Psychologists studying individual differences in intelligence have applied the method of factor analysis to explore the structure of abilities responsible

for performance on various mental tests. With the help of factor analysis, the large correlation matrices that consist of the intercorrelations of diverse cognitive tests can be simplified, assuming that the correlation between any two tests is the result of the tests' correlation with a latent variable, which is not directly measurable.

Factors are therefore latent variables, with which the original tests correlate, and we call this correlation the loading of the test on the factor. The correlation between two tests (A and B) equals the product of the test's loading on the factor (F):

$$r = (A \cdot B) = r(A \cdot F) \times r(B \cdot F)$$

The idea behind factors is that they (a) explain correlations among tests, and (b) can reproduce the original correlation matrix based on factor loadings. With the invention of confirmatory factor analytic methods, it has become possible to test elaborate causal models of latent and manifest variables, similarly, by evaluating how much a hypothetical covariance matrix, derived from the hypothesized links between latent and manifest variables, fits the actual, empirical covariance matrix (in order to fully grasp the logic of latent variable analysis, we suggest readers unfamiliar with this topic refer to a brief overview presented in the [Appendix](#)).

The early development of factor analysis, as well as the first systematic study of individual differences in cognition owe a great deal to Charles Spearman. Spearman discovered that if one examines a matrix of correlations between various mental tests, the most apparent thing one notices is that all the correlations are positive. This is called the “*positive manifold*”, and according to [Jensen \(1998\)](#) it is the most replicated result in all Psychology. Spearman wanted to establish that all of these positive correlations could be accounted for by postulating a single source of variance. In order to do so, he invented an early precursor of factor analysis, the so-called “method of tetrad differences”, with which provided an objective empirical test of general ability theory.

2.2. A Single General Factor

When [Spearman \(1904\)](#) applied this method, he found that the correlations could indeed be accounted for by a general factor, which he termed *g*. The initial model that he developed is often referred to as a two-factor theory, but it is actually consistent with the one-factor solution depicted in [Figure 7.1](#). For Spearman, the two “factors” referred to (a) a general factor, *g*, and (b) several specific factors, called *s*, one for each individual manifest variable. In [Figure 7.1](#), the task-specific abilities are contained in the “error” terms. A more appropriate

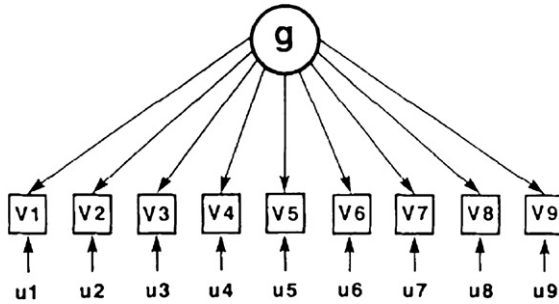


Figure 7.1 A latent variable model consistent with Spearman’s original model of intelligence: variance in each measured variable (V_1, V_2 , etc.) can be decomposed to its loading on the general factor, g , and its unique source of variance (u_1, u_2 , etc., equivalent to Spearman’s s factors). The correlation between the variables is entirely the function of their loading on the general factor. (From Jensen and Weng (1994), p. 243).

interpretation of these latent variables is to consider part of their influence systematic, i.e. specific to the particular manifest variable, and part of their influence as unsystematic, i.e. measurement error. Of theoretical importance is Spearman’s insistence on the s factors being strictly unique to the given test; there is no common variance between tests that is not explained by g , that is, there are no “kinds” of tests that belong to different (verbal, spatial, etc.) abilities.

2.3. Several Group Factors

Thurstone, the sharpest contemporary critique of Spearman’s model, was more concerned with patterns of convergence and divergence than he was with the positive manifold. He argued that two tests within a particular domain (e.g. verbal ability) tend to be more highly correlated with each other than two tests from different domains (e.g. verbal and spatial ability), despite the fact that the correlations among all tasks are positive. Thurstone (1938) questioned the necessity of a general factor and argued for a model of intelligence that included seven primary mental abilities and did not include a general factor (Figure 7.2).

Thurstone applied a slightly different statistical method than Spearman. He did not try to maximize the variance explained by a single factor, but rather looked for factors that were independent of one another, each of which explained a sub-set of the correlation matrix, and together explained as much of the entire variance as possible. The independence of factors, achieved through so-called factor rotation, meant that each test had a loading on a single factor only (this is called simple structure in the terminology of factor analysis).

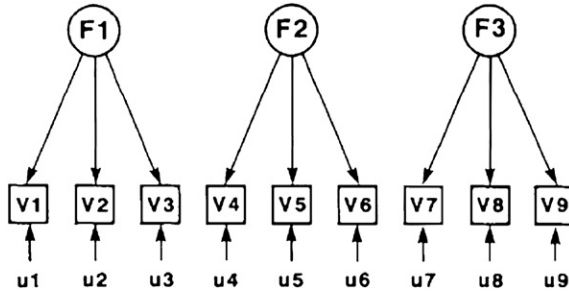


Figure 7.2 A latent variable model consistent with Thurstone’s model of intelligence, displayed here with three rather than seven factors: variance in each measured variable (V1, V2, etc.) can be decomposed to its loading on one of the specific ability factors (F1, F2, etc., representing ability domains, such as verbal, spatial, etc.) and its unique source of variance (u1, u2, etc.). The correlation between the variables is entirely the function of their loading on the general factor. (From Jensen and Weng (1994), p. 243).

The technique of rotation of factors, which is an umbrella term comprising of several actual statistical techniques, made it possible to provide different factorial solutions to any given correlation matrix. Kline (1991) concisely summarizes why this is problematic: “it is (...) clear that the rotated factors may take up any position in factor space and that accordingly, as has been argued, there is a virtual infinity of solutions. Since, as has been seen, these are mathematically equivalent there is no mathematical reason for choosing one rather than another” (p. 61).

Opponents of the psychometric measurement of intelligence, most notably Gould (1996) used this fact to argue that factors are mere mathematical abstractions, and they cannot be meaningfully grounded in differences in actual cognitive processes or neurological attributes. The point of the argument is that since there are an *infinite* number of factorial solutions, it is not possible to reject any of them. This argument, however, rests on a logical fallacy: in reality, the fact that there are an *infinite* number of mathematically equivalent solutions does not imply that *any* factorial solution will be acceptable. For instance, the set of natural numbers consists of an infinite amount of numbers, yet neither “-1” nor “0.5” are parts of the set of natural numbers. Similarly, even if there are an infinite number of mathematically equivalent factorial solutions, it is still possible to show that a given factorial solution is incorrect, and that is exactly what happened with Spearman’s and Thurstone’s original models: in the light of large-scale data sets they turned out to be lacking in terms of model fit.

A single factor proved to be insufficient to explain all of the variance, because there are kinds of tests (e.g. spatial, verbal, etc.) that correlate more with one another than with other kinds. The view that each test loads onto a single group factor has also proved to be untenable. Evidence supports the existence of both a general factor and group factors: a large proportion of the total variance can be explained by a single factor, but there are tests that correlate more with one another than with other tests, (for instance, vocabulary and reading comprehension correlate more with one another than with mental rotation), and such tests load on the same group factor.

2.4. Higher-order and Bi-factor Models

“Higher-order” factor models have been proposed to incorporate both a general factor and multiple second-order factors. These factors are arranged in a hierarchical manner, for instance, in a second-order factor model the first-order (i.e. group) factors correlate with one another, and the general factor emerges because it explains the covariance of the first-order factors the same way as the first-order factors explain the covariance of the manifest variables. Several models of this type have been proposed. Carroll (1993a) summarized a huge number of exploratory-factor analytic studies, over 460 data sets, and provided a single, systematic framework, synthesizing a century of research on the structure of individual differences in mental abilities. Carroll’s three-stratum model, which is presented in Figure 7.3, acknowledges both *g* and specific factors at different levels of a hierarchy.

In a higher-order factor model, such as Carroll’s, the general factor depends on first- and second-order factor loadings. This is not a problem if model fit is the only concern but it is a serious constraint if latent variables are to be interpreted in terms of psychological processes, since it means that the mechanisms responsible for the general factor do not have a direct effect on test performance, rather, they must be some common parameters effecting all abilities (such as speed, etc.). An alternative to the hierarchical solution is the bi-factor model (originally proposed by Holzinger & Swineford, 1937; Figure 7.4), in which each variable loads on both a group factor and a general factor, which means that the correlation between the general factor and an individual variable is not constrained by the correlation with the group factor. Moreover, it has been shown that the hierarchical and the bi-factor solutions are mathematically equivalent or near-equivalent (Mulaik & Quartetti, 1997). However, these models are compatible with different accounts of within-individual mechanisms in general, and different theoretical interpretations of the general factor in particular.

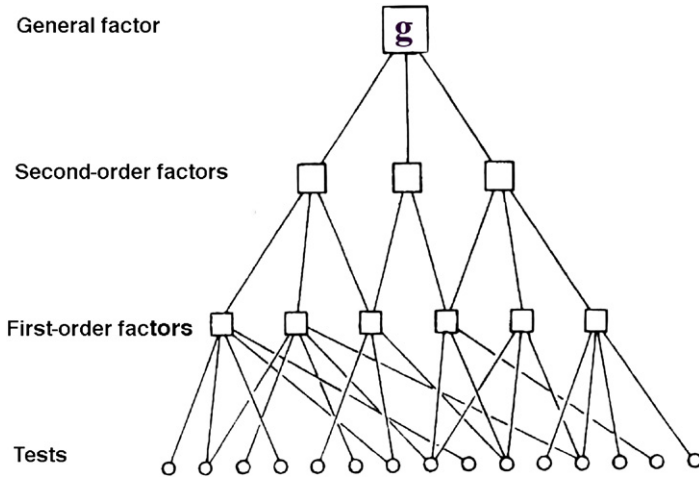


Figure 7.3 Carroll's hierarchical latent variable model of intelligence: tests load on first order factors, which load on second order factors, which load on the general factor.

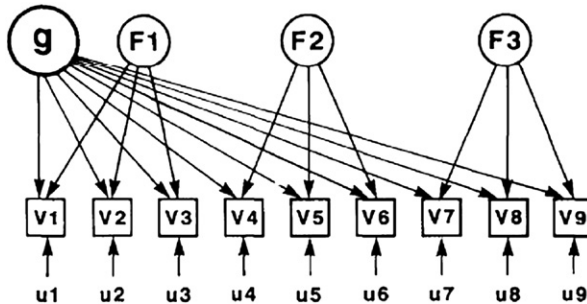


Figure 7.4 A bi-factor latent variable model of intelligence: the variance in the test scores (V_1 , V_2 , etc.) is the function of their loading on the general factor (g) and one of the specific factors (F_1 , F_2 , etc.), with the additional unique variance of each test (u_1 , u_2 , etc.). (From Jensen and Weng (1994), p. 245).

Finally, the radex model proposed by Snow, Kyllonen, and Marshalek (1984) arranges factors by two-dimensional scaling according to complexity, with the most complex tasks located in the middle of the radex. The individual points on Figure 7.5 represent tests and the distance between points is inversely related to the correlation between the two tests, i.e. the further apart the points the smaller the correlation between them. Once again, the radex model has been shown to be parallel to the hierarchical solution.

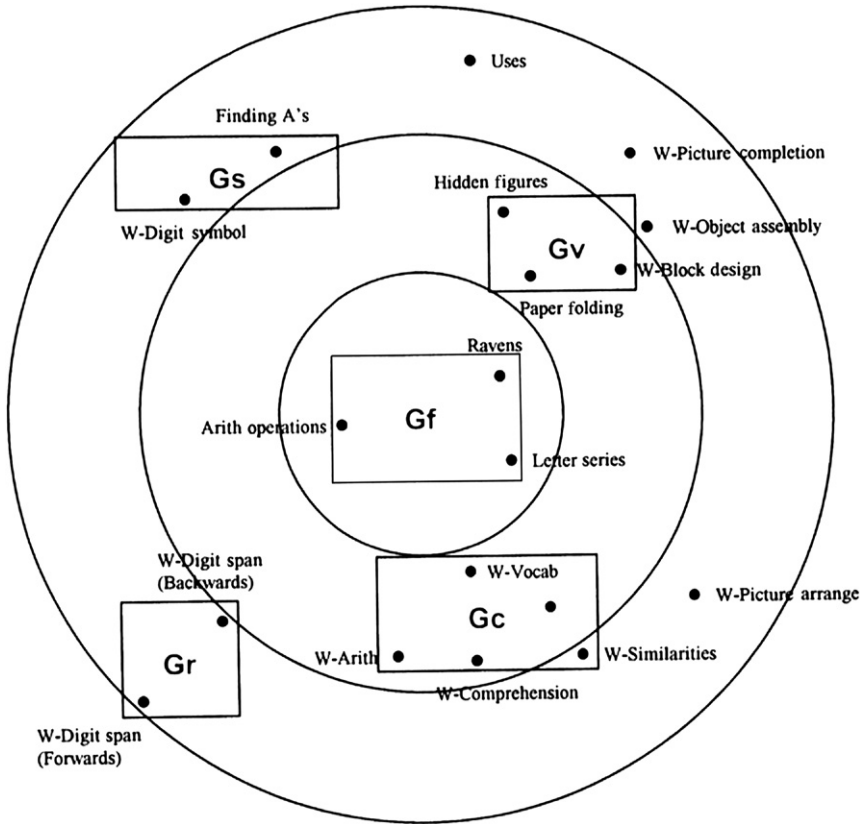


Figure 7.5 The radex model of intelligence (from Mackintosh, 2011, p. 53.): the dots represent individual tests, and the distance between any two tests is the inverse of their correlation. This figure is a greatly simplified version of Snow et al.'s (1984) original.

2.5. The Content of Factors

The three-stratum, the bi-factor, and the radex models are all adequate structural descriptions of the positive manifold, and the convergence and divergence found in the patterns of correlations. But what about the content of the factors?

The most influential account of factor content is the fluid–crystallized model proposed by Cattell (1971); Horn (1994). There are several primary abilities in their model, but the main idea, which will have great significance later on, is the distinction between the ability to solve problems in novel situations, regardless of previously acquired knowledge (fluid intelligence or *Gf*) and the ability to solve problems using already acquired skills or knowledge (crystallized intelligence or *Gc*). The other most important

factors in the model are Gv (visual–spatial), Gs (speed) and Gr (retrieval from memory). (Figure 7.5 for examples of typical measures of these factors). It is interesting to note that in the radex model, Gf, the factor of fluid intelligence is in the middle of the figure, which means that the correlation of Gf with other factor is stronger than the other factors' correlation with each other.

Johnson and Bouchard (2005) argued that a fundamental flaw of the fluid–crystallized model is that it denies the existence of a general factor on the grounds that general factors extracted from different batteries are not the same. However, large-scale analyses show that this is not necessarily the case, and that general factors may be identical across batteries (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004, 2008). Instead of Gf–Gc, they propose their own model: verbal–perceptual–mental–rotation, which is basically an extension of Vernon's model (Vernon, 1961). Unlike Cattell and Horn, Vernon did postulate a general factor, along with two broad second-order factors, v:ed for verbal–educational abilities and k:m for kinesthetic and mechanic abilities. The group factors are much more content-oriented: whereas Gf and Gc are basically described by whether one has to deal with novel or already learned information, v:ed and k:m are described by the domain they cover.

In the last 15 years, a new structural account of intelligence has emerged: the CHC (Cattell–Horn–Carroll) model (McGrew, 2009), which merges the fluid–crystallized model with Carroll's three-stratum account. Since there is a large similarity in the factors of the Gf–Gc model and factors in the second strata of Carroll's model, the merge is relatively easy, apart from one crucial difference: the former does not allow a general factor, whereas on stratum three there is *g*. That is, while almost completely accepting the factorial structure of the Gf–Gc model, CHC does allow a third-stratum general factor to explain the covariance between factors at the second stratum. Therefore, with regard to the format of the model, CHC is the same as Carroll's model (Figure 7.3), whereas the second-order factors in the model follow Gf–Gc factors almost completely.



3. INTERPRETATION OF THE GENERAL FACTOR

3.1. A Unitary Source

From a psychometric point of view it might be sufficient to explore the structure of variation in cognitive abilities, but from a psychological perspective an explanation rather than a pure exploration of individual differences

is warranted. Nevertheless, latent variable models historically have emphasized model fit, and have been difficult to interpret in terms of psychological theory. As emphasized in the Introduction, one should be cautious when proposing intraindividual processes as possible candidates of the sources of interindividual differences, and unfortunately most theorists have not proceeded with such caution.

Since the positive manifold is the most ubiquitous phenomenon in the field of cognitive abilities, most explanations have tried to explain why all mental tests correlate positively with one another. Given that the general factor is nothing but a statistical equivalent of the—general—positive manifold, such explanations have taken the form of theories of g .

The first explanation of g was provided by its discoverer, Charles Spearman. He believed that the general factor is the factor of *general intelligence*, and described it as a general mental energy. He applied the analogy of engines: the specific abilities (s factors) are thought of as independently operating engines, while the general factor is analogous to the source of power that supplies all the engines. Therefore, he postulated the existence of a domain-general cognitive process.

This is certainly a parsimonious explanation: it is perfectly reasonable, unless evidence to the contrary exists, that performance on two tests correlates because they tap the same psychological process. However, it is certainly not the only explanation either, and it is important to discuss in the light of available evidence, whether g is unitary or not. But, what kind of evidence should be taken into account?

Kranzler and Jensen (1991a, 1991b, 1993) and Carroll (1991a, 1991b, 1993b) had a series of discussions about the unitary nature of the general factor. Kranzler and Jensen factor-analyzed various so-called elementary cognitive tasks (such as various reaction time measures) and found different “elementary cognitive factors”, many of which correlated with the g factor extracted from psychometric tests, but not with each other, which lead them to conclude that g is the result of several independent processes. Carroll debated this explanation and claimed that the procedure used by Kranzler and Jensen could not extract pure factor scores, hence could not yield a clear picture of the subjects’ level of g . He did not claim that g is unitary, he only argued that the question could not be decided by the methods employed by Kranzler and Jensen.

Jensen summarizes the debate claiming: “if successful performance on every complex mental test involves, let us say, two distinct, uncorrelated processes, A and B (...) in addition to any other processes that are specific to

each test or common only to certain groups of tests, then in a factor analysis all tests containing A and B will be loaded on a general factor. At this level of analysis, this general factor will forever appear unitary, although it is actually the result of two separate processes, A and B” (Jensen, 1998, p. 261.) This is perfectly in agreement with what has been said in the Introduction about the under-determination of individual differences data to theories of intraindividual process.

However, Jensen takes one step further: “to show that the general factor involves individual differences in two independent processes, A and B, and is therefore not fundamentally unitary would require that individual differences in A and B be measured separately and that A and B are each independently correlated with the general factor of the psychometric tests. The more difficult condition to satisfy (...) is that it must be assumed that the empirical g factor scores derived from the tests are “pure” g uncontaminated by any non-g “impurities”. (...) [But] because it is virtually impossible to prove definitively that the g factor scores are “pure” in this sense, *the issue retreats from the scientific arena, and it then becomes a purely metaphysical question whether g is or is not unitary*” (Jensen, 1998, p. 261, italics added).

This is a surprising conclusion. Arguably, it is not metaphysics, but rather experimental psychology and neuroscience that should be able to answer whether g can be equated with a unitary intraindividual process, since these are exactly the disciplines that have the methods to fractionate both cognitive and neural mechanisms. Experimental psychologists generally claim that two processes are different if the tasks that purportedly tap the two processes respond differently to changes in experimental conditions. In neuropsychology, two processes are dissociated if injury or developmental impairment in one part of the brain disrupts one process while leaving the other intact.

3.2. Evidence Against a Unitary Source

There is now ample evidence from experimental cognitive psychology, neuropsychology, and neuroscience, which directly contradict the idea of a domain-general problem-solving mechanism involved in all kinds of cognitive activity. Experimental studies have demonstrated the fractionation of abilities, primarily verbal and spatial. For instance, experimental manipulations that engage the cognitive system responsible for the short term storage of verbal information impair performance in verbal, but not spatial reasoning tasks, whereas a similar engagement of the short-term spatial store results in impaired performance in spatial reasoning while verbal reasoning remains intact: this result was crucial

for proposing domain-specific short-term stores in the working memory model (Baddeley, 1992). Some genetic disorders resulting in cognitive malfunction also selectively impair spatial and verbal abilities (e.g. Williams syndrome and specific language impairment, respectively). Moreover, localized damage to frontal brain regions can cause impairment on tests measuring G_f , such as those of nonverbal reasoning, while performance on tests of G_c , such as those measuring vocabulary and general information, remains intact (Duncan, 1995; Duncan, Burgess, & Emslie, 1995).

Also, the secular increase in IQ-test results, the so-called Flynn-effect, is observed to varying degrees across different tests. Increases in average raw scores in most ability tests have been observed in the last century (Flynn, 1987). However, on tests of abstract inductive reasoning (G_f) the gains equal about 15 IQ-points per generation, whereas on tests of general knowledge, arithmetic, or vocabulary (G_c), the increases are only marginally significant, in the range of 3–5 points per generation.

Finally, evidence of sex differences also supports the fractionation of g into components: the general tendency for males to excel in spatial tests, especially mental rotation, and for females to excel in various tests of verbal ability show that even though performance on these tests correlate, they cannot measure a single ability. If they did, and verbal and spatial tests measured the same general intelligence, then males or females should be better on both verbal and spatial tests (Mackintosh, 1996). All these different lines of evidence converge toward the conclusion that g cannot be equated with “general intelligence” as an individual-level mechanism, such as one that makes it meaningful to say: “John used his general intelligence to solve test items in both the vocabulary test and the one requiring mental rotation.”

3.3. Multiple Sources

There have been a number of attempts to explain the general factor (that is, the positive manifold) without reference to a general intelligence. In 1916, Godfrey Thomson illustrated that a unitary source of variance is not necessary to account for the presence of a general factor. That is, unlike Thurstone, he admitted the existence of the general factor, but denied that it could be identified with a single, unitary process. Thomson proved mathematically that a general factor can emerge not just without a single general underlying mechanism, but even without a single process being common to all of the tests. He argued that the positive manifold manifests itself because any battery of mental tasks will “sample” processes (which he called ‘bonds’) in an overlapping manner, such that some processes will be required by a shared

subset of tasks, while other processes will be unique to particular tasks. Two decades later, Tryon (1935) proposed a similar, but much more ambitious model, and argued that unitary-source models would eventually become untenable with respect to psychological processes (he was ahead of his time).

The idea of overlapping psychological processes is elaborated by Mackintosh (2011). His explanation is illustrated in Table 7.1. Suppose a battery of six tests was administered to a group of subjects and across the battery, nine psychological processes were sampled. Tests that share several processes will form a cluster, suggesting group factors, but the fact that every test shares at least one process with another test will result in a general factor.

Using random data, Thomson was able to show that a general factor can be explained both by postulating a single general ability or a large number of independent processes. There is another important question: whether there are *any* processes common to all mental tests, or whether the process common to tests A and B is quite different from the process common to A and C, which is in turn different from the process common to B and C. Thomson also showed that this last possibility is logically just as consistent with the general factor as the alternative answer that there is one, or there are several, processes common to all tests.

Thompson's model was originally referred to as a "sampling theory" but it has been recently revised and updated as "the bonds model of intelligence" (Bartholomew, Deary, & Lawn, 2009). Bartholomew et al. also showed that there is no mathematical way of choosing between Spearman's and Thomson's models, since they both provide plausible accounts of the positive manifold: once again, it has been demonstrated that theories of intraindividual cognitive processes are underdetermined by individual differences data.

Table 7.1 Nine Hypothetical Processes

Six Tests	1	2	3	4	5	6	7	8	9
Gc									
1	+	+	+	+					
2		+	+	+	+				
Gf									
1	+			+	+	+		+	+
2	+		+		+	+	+		+
Gv									
1		+				+	+	+	
2		+					+	+	+

Finally, the “mutualism” model of g by van der Maas et al. (2006) proposes that the positive manifold arises because independent cognitive processes engage in mutually beneficial interactions during cognitive development. Thus by the end of development the processes will be correlated, resulting in the positive manifold, but there is no single ability underlying performance on all tests.

In summary, the general factor of intelligence can be, and has been, interpreted in many ways. Which interpretation is most valid? If one were to consider evidence solely from research on interindividual differences then it is impossible to judge. However, if we consider converging evidence, for example, from neuroscience, and if we consider models of intraindividual processes, in particular models of working memory, then it may be possible to favor one interpretation over others. In our view, theories of general intelligence that posit a unitary source of variance should be rejected. As discussed, several new areas of research, especially in neuroscience, simply do not support such models. In our opinion, the rejection of general ability theories of intelligence, on the basis of solid empirical evidence rather than conjecture or politics (e.g. Murdoch, 2007), may eventually be viewed as one of Psychology’s greatest achievements.



4. MODELS OF WORKING MEMORY CAPACITY

4.1. History Repeating

The “positive manifold” is not confined to the correlation matrices of the intelligence literature. The same type of positive manifold is commonly observed when a battery of working memory tasks is administered to a large group of subjects. As with batteries of intelligence tests, patterns of convergence and divergence are typically observed amidst the positive manifold. For example, working memory tasks with verbal content tend to be more strongly correlated with other verbal tasks than with tasks with spatial content. Yet the positive manifold is still observed.

As mentioned in the Introduction, the first “test” of working memory capacity was the reading span task (Daneman & Carpenter, 1980). The task requires subjects to read sentences aloud and remember the last word of each sentence for later recall. The number of sentences/words per list varies, typically from two to six or seven. Daneman and Carpenter found that working memory capacity, as measured by their reading span task, was strongly correlated with Scholastic Aptitude Test (SAT) scores. This is not a surprising result, given that reading span and the SAT, particularly the verbal sections,

both require *reading*, and so the correlation could be interpreted to reflect domain-specific variability. Indeed, Daneman and Carpenter suggested that domain-specific processes largely determine working memory capacity.

An alternative interpretation is to assume that domain-general processes largely drive working memory capacity. Turner and Engle (1989) were the first to offer evidence in support of a domain-general view. They demonstrated that a complex span task that does not require reading (the operation span task) predicted SAT scores just as strongly as reading span, and moreover, operation span and reading span accounted for the same variance in SAT scores.

In effect, Turner and Engle (1989) sparked a debate about generality vs. specificity with respect to working memory capacity. Yet it is all just a little bit of history repeating. The debate bears a striking resemblance to the debate between Spearman and Thurstone. On one side is the more general/unitary view, which assumes that variation is largely caused by domain-general factors and on the other side is the specificity view, which assumes that variation is largely caused by more specific factors. In the end, the two sides acknowledge the existence of *both* domain-general and domain-specific sources of variation but they argue about their relative importance. The structure of working memory capacity is therefore analogous to the structure of intelligence and can be explained by the same latent variable models that were used to explain the structure of intelligence (specifically, the hierarchical and bi-factor models).

As an illustration, consider two models of working memory capacity published by Kane et al. (2004), shown here in Figures 7.6 and 7.7. Kane et al. administered several memory span tasks to a large heterogeneous sample of subjects. Some tasks were simple span and others were complex span, and some were verbal and others were spatial. The model in Figure 7.6 depicts a unitary, domain-general capacity that is strongly correlated with fluid intelligence. Kane et al. argued that this one-factor model of capacity fits the data as well as a two-factor model and therefore chose the more parsimonious one-factor model. The fact remains, however, that the two-factor model fits too! So again, it is possible to account for interindividual differences with alternate models. The model in Figure 7.7 is similar in structure to the bi-factor model of intelligence and demonstrates that whatever complex span tasks have in common, after accounting for simple span, is strongly correlated with fluid intelligence. Kane et al. interpreted this common variance to reflect “executive attention.” We discuss this interpretation, and others, in the next section.

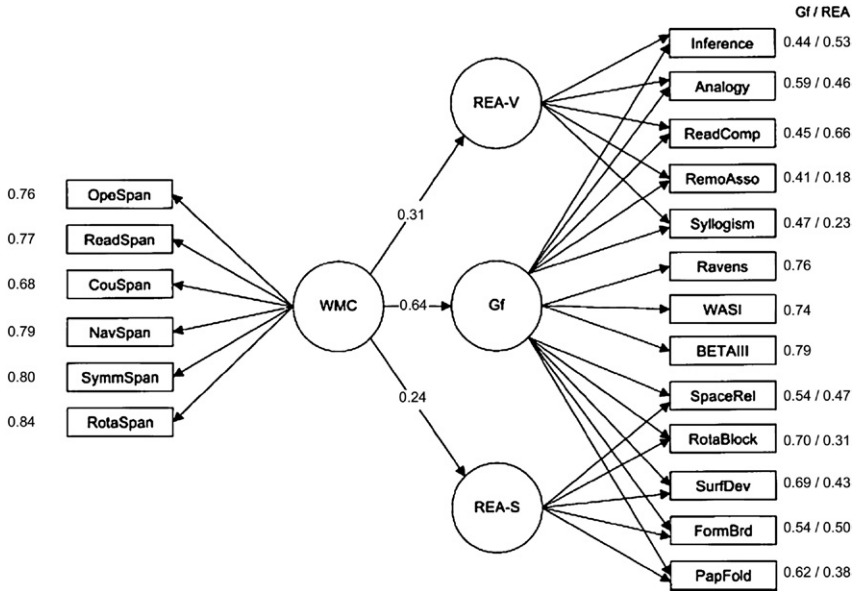


Figure 7.6 Kane et al.'s (2004) one-factor latent variable model of working memory capacity and its relation to general fluid intelligence (Gf).

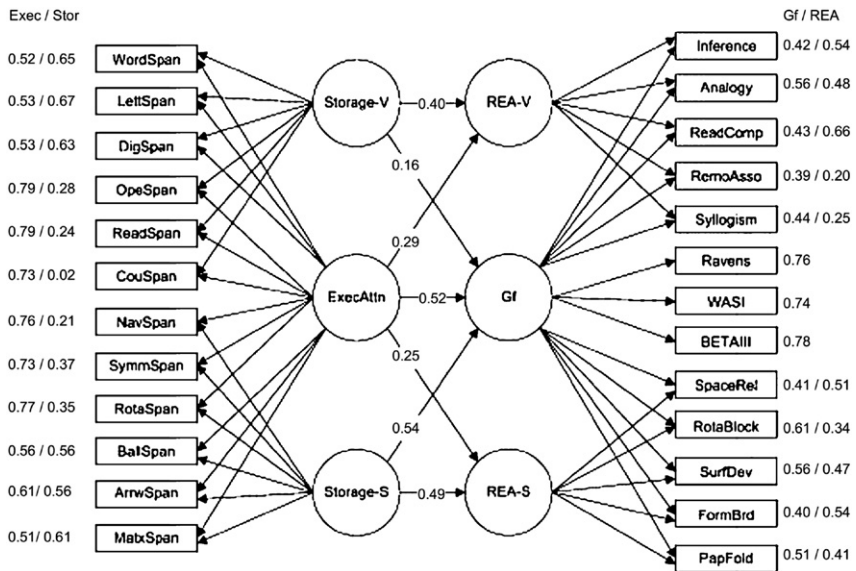


Figure 7.7 Kane et al.'s (2004) "executive attention" interpretation as the general factor underlying working memory task performance.

4.2. The General Factor, Again

The models published by Kane et al. (2004) are obviously not the only latent variable models of working memory capacity to consider but their theoretical framework is among the most influential in the field of working memory. As well, their “executive attention” view of individual differences in working memory capacity is supported by a wealth of data from both correlational and experimental psychology (Engle and Kane, 2004). This particular framework therefore uniquely informs our current review. In short, there are pros and cons to their unified approach. We discussed the many benefits of a unified approach in the Introduction so here we consider the challenges.

Perhaps the most difficult challenge for unified approaches, and a challenge for executive attention theory, is articulating the distinction between the model that accounts for interindividual differences and the theory of working memory. For example, executive attention theory is widely considered a unitary model because Kane et al. (2004) favored a one-factor model of working memory capacity. This does not imply, however, that they also endorse a unitary model of working memory in terms of intraindividual psychological processes. Unfortunately, unified theories are so uncommon in Psychology that this subtle point is lost on many researchers, so the onus is on the authors to repeatedly clarify their position. To their credit, Kane et al. (2004) were clear on this point. After endorsing the unitary latent variable model, they explicitly stated that multiple mechanisms/functions/processes contribute to working memory task performance. They explain “executive attention” as follows: “we believe that these executive-attention functions reflect a capability to maintain information in an active, easily accessible state, whether that information corresponds to a list of several external stimuli or a single goal for action. Moreover, such active maintenance is particularly important in the presence of interference, which disrupts rapid retrieval of information from long-term memory (LTM) and in the blocking of distraction and competing response tendencies.” (p. 213.) The plural phrase, “executive attention functions”, clearly implies that “executive attention” is nonunitary in terms of intraindividual psychological processes.

More recently, Unsworth and Engle (2007) revised and extended the executive attention view. Their theory presents an even more dramatic shift away from a unitary/general ability account because it is nonunitary BOTH in terms of interindividual differences and in terms of intraindividual psychological processes. According to their view, multiple domain-general

mechanisms are engaged in task performance AND contribute variation to task performance. Specifically, they argue that working memory capacity is determined by mechanisms required for active maintenance of information as well as mechanisms required for controlled, rapid search of long-term memory. Active maintenance and controlled search are operationally defined in distinct ways, they account for unique variance in intelligence, and neuroimaging research suggests that they depend upon distinct neural pathways (Chein, Moore, & Conway, 2011; Unsworth, Spillers, & Brewer, 2011). The emerging view is that there are multiple and independent psychological processes involved in the performance of working memory tasks, and multiple and independent sources of variance contributing variation in task performance. And critically, the psychological processes and variation in behavior can be linked in a theoretically meaningful fashion (e.g. Unsworth & Engle, 2007).

In sum, the general factor in models of working memory capacity does not appear to be linked to a single psychological process. We prefer to interpret the general factor to reflect multiple domain-general mechanisms that are tapped in an overlapping fashion across a battery of working memory span tasks. This is consistent with our view of the general factor in models of intelligence, and supports our contention that unitary source models of general cognitive ability are no longer viable and new lines of research should be encouraged to further investigate nonunitary, multi-mechanism models.



5. TOWARD A UNIFIED MODEL

5.1. Intelligence and Working Memory Capacity

Having independently reviewed both intelligence and working memory, it is time to turn our attention to the link between these constructs. The closest possible link between any two constructs, of course, is identity. Are intelligence and working memory identical constructs? Or, more precisely, are individual differences in intelligence and individual differences in working memory capacity caused by variation in the same processes?

As we have seen, a positive manifold is observed among measures of working memory capacity as well as measures of intelligence. Two decades ago, Kyllonen and Christal (1990) published a series of studies suggesting that psychometric working memory capacity might be equivalent to psychometric *g*. While some authors continue to claim that working memory capacity and *g* are isomorphic (Colom et al., 2008), two meta-analyses

concluded that the two constructs are strongly correlated but not quite equivalent (Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Süß, 2005).

What causes, then, working memory capacity and g to share a large part of their variance? This question can be decomposed into two different ones: (a) which component of working memory capacity is responsible for its correlation with g , (b) which component of g is responsible for its correlation with working memory capacity? Let us answer these questions in order.

5.2. Components of Working Memory

To answer the first question we must consider all the components of working memory and assess the degree to which each component influences task performance. Engle and Kane (2004) provided a detailed illustration of the components of working memory (Figure 7.8). This illustration is particularly useful here because it allows us to succinctly explain three important points. First, complex span tasks are considered to be more accurate measures of working memory capacity than simple span tasks. Simple span tasks, such as digit span and word span, do not involve a secondary processing task

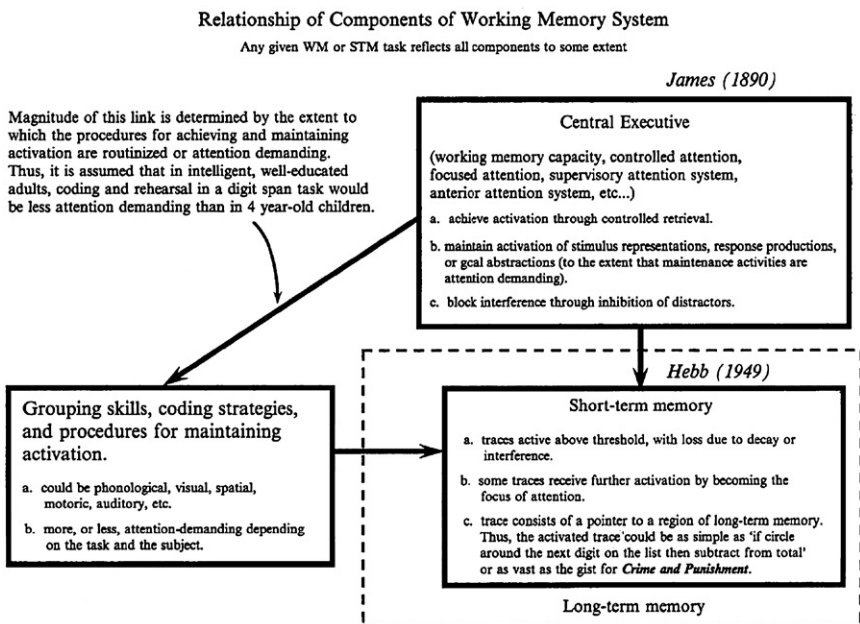


Figure 7.8 Engle and Kane's (2004) illustration of the components of working memory.

(such as reading sentences aloud in reading span). Simple span performance is therefore considered to be more strongly influenced by the components listed in the left-most box in [Figure 7.8](#). In contrast, complex span task performance is thought to be more strongly influenced by the components listed in the top box, labeled Central Executive.

The second point is the distinction between domain-general and domain-specific influences on task performance. The Central Executive component is considered to be more domain-general and the Skills/Strategies component is considered to be more domain-specific. Finally, the third point is that multiple domain-general mechanisms are listed under the umbrella term Central Executive. Again, this suggests that multi-mechanism views are certainly more viable at the level of intraindividual differences and have potential at the level of interindividual differences as well.

Taken together, this framework suggests that the components of working memory that contribute to the correlation with g are associated with the central executive and are more directly tapped by complex span tasks than simple span tasks. For brevity, we refer to these components collectively as executive/attentional mechanisms.

5.3. Components of Intelligence

Now let us turn to the question of what causes variation in g to be in large part common with variation in working memory capacity. As mentioned, a good model of intelligence and working memory should prohibit certain correlations while predicting others. The Gf - Gc model passes this test: it predicts that the executive attention component of working memory, that is, what complex span tasks measure beyond simple storage and retrieval should correlate with tests of fluid (Gf), but not, or to a much smaller extent, with tests of crystallized intelligence (Gc). Two large-scale studies verify this prediction.

The first one is the meta-analysis of working memory and intelligence by [Ackerman, Beier, and Boyle \(2005\)](#), which, although independently explored short-term memory's and working memory's correlation with various types of cognitive tests, did not originally compare these results. The following figure ([Figure 7.9](#)) shows in decreasing order the difference in correlations with working memory and short-term memory in different types of ability tests (based on [Kovacs, 2009](#)).

It is clear that on one side, with the largest difference, are the Raven's Progressive Matrices (Gf), whereas on the other side, with negligible differences, are tests of general knowledge as well as the ones with verbal content (Gc), and tests measuring perceptual speed (Gs). In the middle, with

significant, but less substantial differences than in the case of Gf, are spatial tests (Gv) and ones that support to measure “general ability” or *g*. Therefore, this result shows that the processes working memory (complex span) tasks tap beyond simple storage are strongly associated with Gf, but not associated, or to a negligible extent, with Gc and Gs.¹

In a study employing partial correlations and analyzing large and representative samples, Kovacs et al. (in press) compared the relationship between forward digit span, backward digit span, and a complex working memory task (Letter–Number Sequencing) on the one hand, and tests of Gf (matrix reasoning) and Gc (vocabulary and general knowledge) on the other. They found that even though a large part of backward span performance can be attributed to the short-term storage of information, what backward span measures above and beyond pure storage and retrieval is closely related to fluid reasoning (Gf) in all age groups, but is only related to crystallized intelligence (Gc) in children aged 6–10. They also found that Letter–Number Sequencing correlates significantly with tests of both Gf and Gc, and generally more strongly than backward span, albeit there is a large difference favoring Gf in this case as well.

Overall, the results considered in this section point to the conclusion that it is the executive/attentional component of working memory and the fluid reasoning component of intelligence that are primarily responsible for the working memory–intelligence relationship.

5.4. Converging Evidence

Another kind of evidence pointing to the possibility of the same processes being involved in working memory and fluid intelligence comes from neuroimaging. Neuroimaging studies of executive processes and working memory highlight the importance of the prefrontal cortex which, in most cases along with posterior parietal areas, is involved in cognitive control (e.g. Miller, 2000), task switching (e.g. Sohn, Ursu, Anderson, Stenger, & Carter, 2000), and inhibition (Aron, Robbins, & Poldrack, 2004). Henson (2001) reviewed neuroimaging studies of working memory and concluded that “the central

¹ Note that in tests of numerical reasoning the difference in explained variance between working memory and short term memory is surprisingly high, almost as much as in the Raven’s Matrices. However, only six tests belong to this category, two of which are number series, and one is “induction-quantitative” (Ackerman et al., 2005, p. 58.). Although these tests do have numerical content, they are clearly measures of Gf. Actually Horn (1989) in his categorization of ability tests according to the Gf–Gc model, put “inductive reasoning, measured using letter series, number series and/or figure series” as the first example of indicators of Gf, “matrices reasoning with visual patterns” comes only second (p. 79.)”.

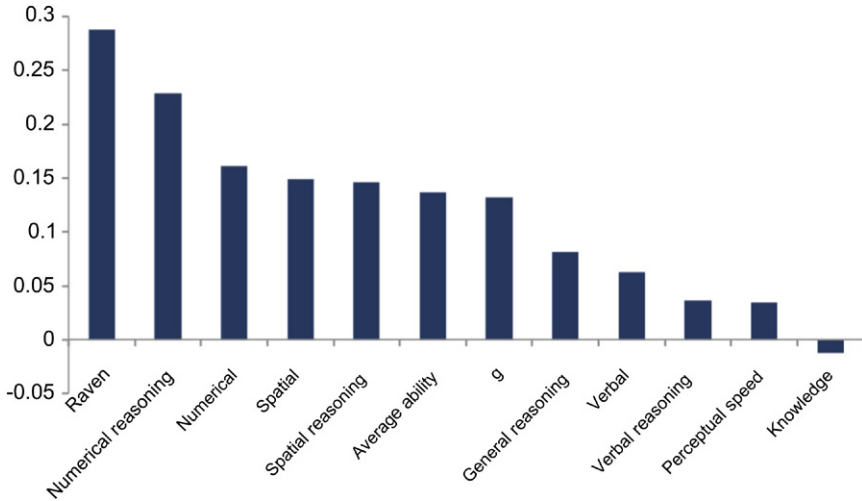


Figure 7.9 The difference between correlations with working memory and short-term memory in different types of ability tests (based on Kovacs, 2009, p. 94.).

executive maps to midlateral prefrontal regions, particularly left and right dorsal lateral prefrontal cortex; the phonological store maps to left inferior parietal cortex; the articulatory control process maps to the left premotor regions (including Broca’s area), left supplementary motor regions and perhaps right cerebellum; the visual cache maps to bilateral anterior occipital and/or inferior temporal regions; and the inner scribe maps to right premotor and right superior parietal regions.” (p. 166–167.).

Wager and Smith (2003) performed a meta-analysis of 60 neuroimaging studies of working memory. They found that the fractionation of working memory according to content type was limited to the posterior areas: they found no evidence of fractionation in the frontal cortex according to content domains. With respect to executive functions, they found that “effects of executive processes were found primarily in the frontal and parietal cortices, including the anterior prefrontal, the dorsal lateral prefrontal cortex (DLPFC) and ventral lateral prefrontal, the bilateral premotor (centered in the SFS [superior frontal sulcus]), and the lateral and medial superior parietal cortices. (...) Only in the aforementioned regions did we find that the presence of executive demand significantly increased the probability of activation” (p. 269–270.).

Similarly, there is ample evidence pointing to the involvement of the prefrontal cortex in fluid reasoning. Duncan et al. (1995) tested three frontal lobe patients with the Wechsler Adult Intelligence Scale (WAIS) and the Cattell Culture Fair test (CCF), and compared their results to healthy

controls matched according to their WAIS scores, which ranged from 126 to 130, and to a group of patients with injury to more posterior regions. They found a large discrepancy between frontal patients' scores on the WAIS and the CCF: their WAIS IQ was 22, 29, and 38 points higher than their CCF IQ. The healthy controls, however, had equal or higher scores on the CCF than on the WAIS. Moreover, when the frontal patients' CCF IQ was compared to that of the healthy controls, the patients' IQ was 23, 51, and 60 points lower. The patients with more posterior damage had lower WAIS scores than frontal patients, but they demonstrated no discrepancy between their WAIS and CCF scores.

Waltz et al. (1999) presented matrix problems to patients, which were similar to the ones in the Raven's Matrices, albeit in a much simpler arrangement. There were three levels of complexity. Level 0, the missing piece of the matrix was identical to all the presented pieces. Level 1 meant a single change in one of the features of the items (e.g. two white squares in the top row, and a black square in the bottom row plus the missing piece), whereas level 2 meant changes in both the horizontal and vertical dimensions (e.g. a white square and a white triangle in the top row, and a black square in the bottom row plus the missing piece). Prefrontal and temporal patients, as well as healthy controls had equal scores on level 0 and level 1 problems (with correct answers ranging between 80 and 100%). The performance of prefrontal patients, however, was seriously impaired on level 2 problems (with approximately 10% correct answers), whereas temporal patients and healthy controls scored at approximately 90% correct on these problems, too.

Neuroimaging studies also confirm that tests of Gf activate the PFC, just like working memory tasks. Duncan et al. (2000) presented items that were created to be similar to the ones in the Cattell's Culture Fair test, and control items that do not load on Gf. They found that the regional blood flow was increased in the PFC for complex Gf-items, but found no similar pattern for items that do not load highly on Gf. Their results support the idea that executive processes, that can be localized in the PFC, primarily in dorsolateral areas, are important in tests of Gf.

In an fMRI study, Prabhakaran, Smith, Desmond, Glover, and Gabrieli (1997) used items similar to the ones that appear in the Raven's Matrices, but categorized them according to whether they required only figural-spatial reasoning or analytic reasoning as well. They found that both kinds of items activated the lateral PFC, but the analytic items activated more prefrontal areas, and brain activity was more bilateral. Moreover, solving analytic items resulted in greater activation than figural-spatial problems. In

a PET study, Wharton et al. (2000) presented figural analogies that required inductive reasoning, and also found activation in the lateral PFC.

Finally, Kane (2005) reviewed two studies that involved the parametric manipulation of complexity in tasks that resembled the Raven's Progressive Matrices (RPM), and concluded that "what seems quite clear from these studies, then, as predicted by working memory (WM)/attention theories of Gf and PFC functioning, is that LPFC regions are recruited in RPM-type problems as their memory load and control demands increase."

In sum, neuroimaging research provides compelling converging evidence to support the conclusion that it is the executive/attentional component of working memory and the fluid reasoning component of intelligence that are primarily responsible for the working memory–intelligence relationship. In our opinion, future work in neuroscience will provide insight into the nature of the relationship between working memory and intelligence because the fine-grained level of analysis will help to clarify the cognitive and neural mechanisms that contribute to task performance (e.g. Burgess, Braver, Conway, & Gray, 2011).



6. CONCLUSION

Having reviewed models of intelligence and working memory, and having specified cognitive processes that might underlie the common variance between these complex systems involved in higher-order cognition, what can we conclude? We present here a brief list of clear and concise arguments. They are presented in bullet-point format because it is our hope that each of these statements has some impact. Consider this the take-home message:

1. Intelligence and WM are related, but not identical constructs, with a large degree of common variance in behavioral outcomes.
2. It is the fluid (Gf) component of intelligence and the executive attention component of WM that is responsible for the relationship.
3. It seems likely that WM capacity (WMC) and Gf are related because some of the processes involved as well as the neural substrates of these processes are identical.
4. It is indeed possible for theories about intraindividual processes to predict the structure of interindividual differences, and the relation between WM and intelligence is a good example.
5. The general factor, *g* or WMC, does not imply a unitary source of variance. This is not novel but the message needs to be delivered loud and

clear because the default interpretation, in science and society, is to infer a general ability. Moreover, the evidence now seems to be AGAINST general ability theories.

6. A unified model of intelligence and WM will have to account for why variance in behavior seems more domain-general than one would expect from multiple individual-level processes.

Hopefully, the recent explosion of empirical research will be soon accompanied by the birth of large-scale theories, which will not only account for the recent findings on the connection between working memory and intelligence, but will also further our understanding of each of these constructs. Such theories will, then, not only help us understand why working memory and intelligence are related, but will also help researchers do justice on competing, yet from an individual differences perspective, equally fit models of working memory and intelligence.



APPENDIX

The purpose of this Appendix is to provide a concise overview of latent variable analysis. It is intended for readers with little prior exposure to the method. For many readers this section is not necessary, this is why it is presented here rather than in the main text. Much of the content for this section is based on more comprehensive reviews, particularly work by David Bartholomew and Denny Borsboom (e.g. Bartholomew, 2011; Borsboom et al., 2003).

Latent variable analysis is an umbrella term and includes both exploratory and confirmatory factor analysis. The latent variable models we consider here are examples of confirmatory factor analysis; a specific model structure is specified, model parameters are estimated, and then the fit of the model is evaluated.

A latent variable model is a statistical model that relates a set of observed variables to a set of latent variables. The observed variables are typically referred to as manifest variables, and we adopt that terminology here. It is assumed that an individual person's observed score on a manifest variable is the result of that person's relative position on the latent variable. For example, if the manifest variable is a working memory span task and the latent variable is working memory capacity then the individual person's capacity is assumed to cause their score on the span task. It is not a necessary feature of latent variable models but most models assume that the relationship between latent and manifest variables is linear and that the distributions

are normal. Finally, it is also assumed that scores on manifest variables are orthogonal to one another after controlling for the latent variables, a property known as local independence.

Consider the two models presented in Figure 7.10. The model on the left depicts the type of model just outlined. According to the model, the latent variable ξ_1 causes outcomes on the manifest variables X_1 , X_2 , and X_3 (causality is depicted with directional arrows). The total variance in a manifest variable consists of the portion explained by the latent variable and the remainder is assigned to an error term, δ . The path coefficients, or factor loadings, λ , quantify the magnitude of the relationship between the latent and manifest variable, and if standardized, take on values between -1 and $+1$. Thus, the proportion of variance in X_1 determined by $\xi_1 = (\lambda_1 \times \lambda_1)$ and the proportion of variance assigned to the error term $= 1 - (\lambda_1 \times \lambda_1)$.

The model on the right is presented to illustrate the importance of causal flow in latent variable models. Students first learning latent variable analysis often make the mistake of drawing the arrows from the manifest to the latent variables. This seems intuitive because the latent variables are not observed and therefore have to be estimated from the manifest variables. However, it is important to remember that a critical aspect of latent variable models, when formulated as scientific theories, is that they make causal claims about the outcomes on manifest variables. Hence the causal flow from latent variables to manifest variables. Of course, it is possible to specify models in which the manifest variables cause the latent variables (this is more common in sociology and economics). In such cases, the latent variable is considered to be endogenous, i.e. the causes of the variable are contained within the model. This is an important distinction, both mathematically and theoretically, which is why it is common to adopt notation that captures the difference (e.g. ξ vs. η).

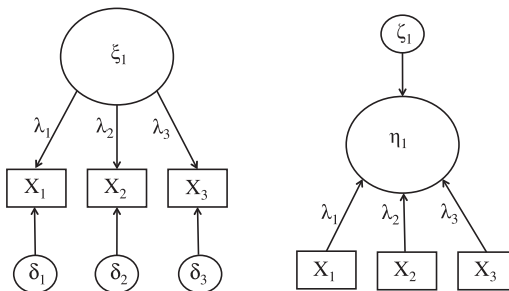


Figure 7.10 Two latent variable models, illustrating different patterns of causality.

In the standard model, there are two ways to think about a latent variable: as a formal-theoretical concept and as an operational-empirical concept (Borsboom et al., 2003). For example, psychometric g may be presented as a theoretical concept, such as general intelligence, or it may simply be presented as an atheoretical factor, often included in a model solely to enhance model fit. In psychometrics, these two interpretations are often referred to as psychological g and psychometric g , respectively. Ideally, the formal and the operational concepts are linked but this is not an empirical process, it is dependent upon one's theoretical perspective.

To connect the formal concept associated with a latent variable with the operational concept, Borsboom et al. (2003) argue that one must adopt a particular ontological stance consistent with the overarching claims of the model, or theory. In their view, there are three positions one can take: operationalism, constructivism, and realism. The operationalist view is the least theoretical; latent variables represent nothing more than empirical content. The constructivist view is also devoid of psychological theory but concedes that the latent variable in question is a construction of the human mind. Realism is the ontological stance we desire here, with respect to g and capacity, in that the construct prescribed to a particular latent variable is thought to exist independent of measurement. This is known as entity realism and is a necessary property of latent variable models that strive to make causal arguments about the relationship between latent and manifest variables.

The “fit” of a latent variable model refers to the correspondence between the observed covariance matrix and the covariance matrix implied by the model. The null hypothesis can therefore be stated as follows:

$$\Sigma = \Sigma(\Theta)$$

where Σ is the population covariance matrix and $\Sigma(\Theta)$ is the covariance matrix implied by the model and Θ is a vector containing the “free” parameters in the model, i.e. the parameters that need to be estimated. There are many indices of model fit but one common method is to compare Σ and $\Sigma(\Theta)$ using a chi-square goodness of fit test and if the chi-square value is not significant then Σ and $\Sigma(\Theta)$ are considered to be equivalent, indicating satisfactory model fit.

Multiple models are often tested on the same set of data. In such cases, the fit of competing models may be compared directly. For example, consider the models depicted in Figures 7.11 and 7.12. The first model suggests a one-factor solution while the second model suggests a two-factor solution. To compare the models, simply calculate the implied covariance model for

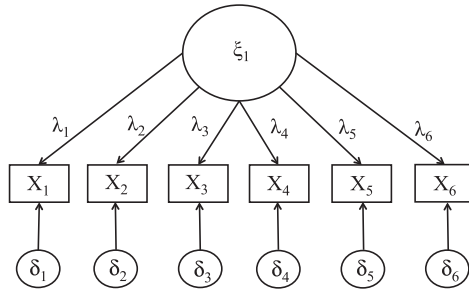


Figure 7.11 One-factor model with six manifest variables.

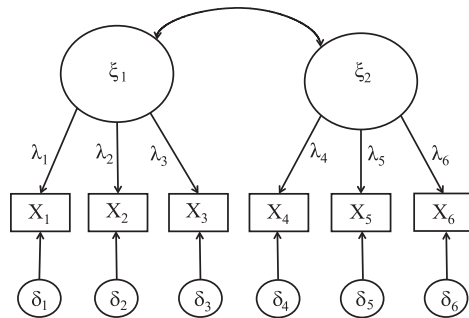


Figure 7.12 Two-factor model with six manifest variables.

each model and then evaluate how much they diverge from each other and from the observed covariance matrix. The expected value for each manifest variable X is estimated via ordinary least squares regression. In matrix form:

$$E(X) = \Lambda_X \xi + \delta$$

where Λ_x is a matrix of factor loadings, ξ is a vector of latent variables, and δ is a vector of residuals. The implied covariance matrix, $\Sigma(\Theta)$, is the product of X and its transpose, X' :

$$\Sigma(\Theta) = E(XX') = [(\Lambda_X \xi + \delta) (\xi' \Lambda_X' + \delta')]$$

In reduced form:

$$\Sigma(\Theta) = \Lambda_X \phi \Lambda_X' + \Theta_\delta$$

where ϕ is the covariance matrix of ξ and Θ_δ is the covariance matrix of δ .

For the one factor model, there are 12 free parameters: Λ_x consists of 5 free parameters (5 factor loadings; 1 factor loading per factor must be fixed), ϕ consists of 1 free parameter (1 variance), and Θ_δ consists of 6 free parameters (6 variances).

In comparison, consider the two-factor model in Figure 7.12. The implied covariance matrix again reduces to:

$$\Sigma(\Theta) = \Lambda_X \phi \Lambda_X' + \Theta_\delta$$

However, now there are 13 free parameters: Λ_X consists of 4 free parameters (4 factor loadings), ϕ consists of 3 free parameters (2 variances, 1 covariance), and Θ_δ consists of 6 free parameters (6 variances).

The point of this exercise is to illustrate that model specification, estimation, and assessment of fit is a rather straight-forward process and provides a standard and objective procedure for evaluating competing models.

REFERENCES

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin*, *131*, 30–60.
- Anderson, M. (1992). *Intelligence and development: A cognitive theory*. Oxford: Blackwell.
- Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*, 170–177.
- Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
- Bartholomew, D. J. (2011). Factor analysis and latent variable modelling. *International Encyclopedia of Statistical Science*, 501–503.
- Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review*, *116*, 567–579.
- Bock, G. (2000). *The nature of intelligence*. Chichester: Wiley.
- Borsboom, D., Kievit, R. A., Cervone, D., & Hood, S. B. (2009). The two disciplines of scientific psychology, or: the disunity of psychology as a working hypothesis. In P. C. M. Molenaar, J. Valsiner, M. C. D. P. Lyra, & N. Chaudhary (Eds.), *Dynamic process methodology in the social and developmental sciences* (pp. 67–97). Dordrecht: Springer.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.
- Burgess, G. C., Braver, T. S., Conway, A. R. A., & Gray, J. R. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*, 674–692.
- Carroll, J. B. (1991a). No demonstration that g is not unitary, but there's more to the story: comment on Kranzler and Jensen. *Intelligence*, *15*, 423–436.
- Carroll, J. B. (1991b). Still no demonstration that g is not unitary: further comment on Kranzler and Jensen. *Intelligence*, *15*, 449–453.
- Carroll, J. B. (1993a). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. (1993b). The unitary g problem once more: on Kranzler and Jensen. *Intelligence*, *17*, 15–16.
- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386–404.
- Cattell, J. M. (1890). Mental tests and their measurement. *Mind*, *15*, 373–380.
- Cattell, R. B. (1971). *Abilities: their structure, growth, and action*. Boston: Houghton Mifflin.

- Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *Neuroimage*, *54*, 550–559.
- Cohen, R. L. (1994). Some thoughts on individual differences and theory construction. *Intelligence*, *18*, 3–13.
- Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*, 584–606.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671–684.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.
- Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.
- Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, *33*, 261–268.
- Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., & Emslie, H. (2000). A neural basis for general intelligence. *Science*, *289*, 457–460.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 44, pp. 145–159). New York: Academic Press.
- Eysenck, H. J. (1995). Can we study intelligence using the experimental method? *Intelligence*, *20*, 217–228.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations—what IQ tests really measure. *Psychological Bulletin*, *101*, 171–191.
- Gould, S. J. (1996). *The mismeasure of man*. New York; London: Norton. Rev., expanded ed..
- Gray, J. R., Chabris, C. E., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*, 316–322.
- Henson, R. N. A. (2001). Neural working memory. In J. Andrade (Ed.), *Working memory in perspective* (pp. 151–173). Hove: Psychology Press.
- Holzinger, K., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, *2*, 41–54.
- Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 61–116). New York: W. H. Freeman.
- Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 1, pp. 443–451). New York: MacMillan.
- Jensen, A. R., & Weng, L. -J. (1994). What is a good g? *Intelligence*, *18*, 231–258.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, Conn: Praeger.
- Jensen, A. R. (2000). “The g factor” is about variance in human abilities, not a cognitive theory of mental structure. *Psycoloquy*, *11*. Retrieved 05/06, 2006, from <http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?11.041>.
- Johnson, W., & Bouchard, J. T. J. (2005). The structure of human intelligence: it is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393–416.
- Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: consistent results from three test batteries. *Intelligence*, *32*, 95–107.
- Johnson, W., Nijenhuis, J. T., & Bouchard, T. J., Jr. (2008). Still just 1 g: consistent results from five test batteries. *Intelligence*, *36*, 81–95.
- Kane, M. J., Tuholski, S. W., Hambrick, D. Z., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217.

- Kane, M. J. (2005). Full frontal fluidity? Looking in on the neuroimaging of reasoning and intelligence. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 141–163). Thousand Oaks, CA: Sage.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2004). *Psychological Bulletin*, *131*, 66–71.
- Kline, P. (1991). *Intelligence: The psychometric view*. London: Routledge.
- Kovacs, K. (2009). *A component process account of the general factor of intelligence*. Unpublished PhD thesis. : University of Cambridge.
- Kovacs, K., Rozsa, S., Ko, N., Meszaros, A., Mlinko, R., Kuncz, E., & Lanyine, E. A. (submitted). Forward and backward digit span measure different components of intelligence.
- Kranzler, J. H., & Jensen, A. R. (1991a). The nature of psychometric g: unitary process or a number of independent processes? *Intelligence*, *15*, 397–422.
- Kranzler, J. H., & Jensen, A. R. (1991b). Unitary g: unquestioned postulate or empirical fact? *Intelligence*, *15*, 437–448.
- Kranzler, J. H., & Jensen, A. R. (1993). Psychometric g is still not unitary after eliminating supposed “impurities”: further comment on Carroll. *Intelligence*, *17*, 11–14.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433.
- Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, *28*, 559–571.
- Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). Oxford: Oxford University Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10.
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, *1*, 59–65.
- Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling: A Multidisciplinary Journal*, *4*, 193–211.
- Murdoch, S. (2007). *IQ: A Smart History of a Failed Idea*. New York: Wiley.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.
- Oberauer, K., Wilhelm, O., Schulze, R., & Süß, H. M. (2005). Working memory and intelligence - Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61–65.
- Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget’s developmental stages. *Acta Psychologica*, *32*, 301–345.
- Popper, K. (1963). *Science as falsification conjectures and refutations*. London: Routledge and Keagan Paul.
- Prabhakaran, V., Smith, J. A.L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D.E. (1997). Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven’s progressive matrices test. *Cognitive Psychology*, *33*, 43–63.
- Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), (Vol. 2). *Advances in the psychology of human intelligence*.
- Sohn, M. H., Ursu, S., Anderson, J. R., Stenger, V. A., & Carter, C. S. (2000). The role of prefrontal cortex and posterior parietal cortex in task switching. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 13448–13453.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, Ill: University of Chicago Press.
- Tryon, R. C. (1935). A theory of psychological components—an alternative to “mathematical factors.” *Psychological Review*, *42*, 425–454.

- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, 30, 128–134.
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1), 104–132.
- Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Inter- and Intra individual variation in immediate free recall: An examination of serial position functions and recall initiation strategies. *Memory*, 19, 67–82.
- van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*, 113, 842–861.
- Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen. Wiley.
- Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cognitive, Affective and Behavioral Neuroscience*, 3, 255–274.
- Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., De Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, 10, 119–125.
- Wharton, C. M., Grafman, J., Flitman, S. S., Hansen, E. K., Brauner, J., Marks, A., et al. (2000). Toward neuroanatomical models of analogy: a positron emission tomography study of analogical mapping. *Cognitive Psychology*, 40, 173–197.

INDEX

Note: Page numbers followed by “f” and “t” indicate figures and tables respectively.

A

ACEs. *See* Action Compatibility Effects

Action Compatibility Effects (ACEs),

96–97, 100–101

for comparing mouse trajectories, 99

mouse tracking task, 103f

Active powers, 169–170

Agent-based computational model

group performance, 19

networks, 22, 24

Particle Swarm Algorithm, 18–19

searching problem space, 22–23

SSEC, 18–19

Algebra problem solving, 128

Algebraic transformations PLM, 127–128

example problem display in, 127f

Ambiguous sketches, 64

Amodal perceptual representations, 141

embodied cognition, 142–144

evidence and argument, 142

neurophysiological evidence, 142

perceptual information, 141

sensory input channels, 141–142

Amodal symbols, 136–137

Ampliation of motion, 182

Anderson’s model, 206–207, 206f

coupling parameter, 209–210

Ahn’s theory, 225

Anterior parietal cortex. *See* Postcentral gyrus

Apprehension Principle, 61

animations, 61, 63

films, 63

interactivity, 63

motion for geometric figures, 62

static graphics, 62

static screen shot, 62f

take-away message, 61–62

Area under the curve (AUC), 99

Aristotle’s force-based approach, 169–170

actions of living things, 170

causal powers theory, 170

natural mechanism, 171

Newton’s theory of force, 172

as psychological causes, 170

Asch’s method, 6

Attention distractors, 89

cell phone, 89–90

GPS systems, 90

in-vehicle technology, 91

navigational aid, 90–91

Attention research, 79. *See also* Memory research

arousal and emotion, 89

arousal-inducing elements, 88

attention in situ, 79–80

automaticity, implications for, 80–84

emotionally arousing information, 88–89

goal processing and, 84–87

levels-of-focus hypothesis, 87–88

AUC. *See* Area under the curve

Automaticity

cognitive, 80–81

inattentional blindness, 80

just-in-time models, 83

location information, 83–84

location processing, 81

location/identity binding trials, 81–82

object identity and location binding, 82f

preattentive feature processing, 81

processing spatial information, 80–81

spatial rehearsal effects, 83

transfer appropriate processing, 84

B

Bayesian inference, 237

Bayesian networks, 205–207

prototype view, 208

Bi-factor models, 244, 246

of intelligence, 253

latent variable model, 245f

Biological motion, 137–138

Biological systems, 118–119

- Bottom-up perceptual information, 50
 Broken ruler problem, 129f
- C**
- Carroll's hierarchical latent variable model, 244, 245f
- Categorization, 204–205. *See also*
 Prediction tests
 causal reasoning, role of, 205
 feature-based accounts, 205–206
 Anderson's model, 206–207, 206f
 exemplar model, 206–210, 208f
 mixture model, 209–210
 prototype model, 206–208, 207f
 models based on causal reasoning
 feature centrality, causal determinants of, 211
 causal model theory, 212
 Murphy and Medin models, 210–211
 theory-based models, 210
- Category-irrelevant contextual features, 221t, 223t
 categorization judgment, 222–223
 effects on categorization, 220
generative causes, 222–224
 participant rates, 220–222
preventative causes, 222–224
- Cattell Culture Fair test (CCF), 260–261
- Cattell–Horn–Carroll model (CHC model), 247
- Causal Bayesian network framework, 213–214, 215f
 category membership, variable representing, 214
 category-irrelevant generative causes, 216f
 experimental design, representation of, 215f
 external causes, 214
 features, 217
 flexibility, 217
 sigmoid function, 216–217
- Causal claims, 168
- Causal illusions, 191–192
- Causal impression effects, 179
- Causal model theory, 212
 of Rehder, 210, 226
- Causal powers theory, 169–170
- Causal reasoning
 causal discounting
 constraint satisfaction models, 224
 discounting prediction, 225
 feedforward networks, 224–225
 PEAT, 225
 causal model theory, 212
 causally based categorization
 Ahn's theory, 225
 Murphy and Medin model, 226
 caveats, 227–228
 feature centrality, causal determinants of, 211
 Murphy and Medin models, 210–211
 conceptual framework, 211
 theory-based models, 210
- Causation, 168
 forces in indirect access, 187–188
 forces in social causation, 188
 causal events, 188–189
- Michotte's arguments
 causal impression, 179
 configuration of forces, 180–181
 forces role in causation perception, 181–183
 for launching effect independence, 179–180
 object properties, 181
 sense of force, 183
 in absence of mechanism, 191–196
 in auditory signal control condition, 186
 causal events, 185–186
 causal vs. noncausal event, 186–187, 186f
 correlations, 189–191
 difference in reaction time, 186–187, 188t
 physical force experiments, 183–185, 184f
 predictions, 190
 trial structure, 185
 in visual signal control condition, 186
- CCF. *See* Cattell Culture Fair test
 CCF IQ, 260–261

- Cell phone, 89–90
- Central Executive component, 258
- CHC model. *See* Cattell–Horn–Carroll model
- Cognitive activity, 143
- Cognitive design principles, 58
- Collective learning
 - drawings, 25
 - fitness functions in experiments, 24
 - round-based picture-matching puzzle game, 25–26, 26f
 - cell-by-cell comparison, 26–27, 27f
 - diversity, 26–27
 - imitation rate, 30–31
 - imitators' similarity comparison, 29, 30f
 - improvement, 26–27
 - results, 27–28, 28f
 - strategy and performance relationship, 28–29, 29f
 - turnover, 26–27
- Collective search in problem space
 - group advantage for individual imitation, 31
 - methodological improvements, 31–32
 - offsetting disadvantage, 31–32
- Complex span tasks, 236
- Conformity, 9–10
 - to other's ideas, 10
- Congruity Principle, 61
 - animations, 61, 63
 - films, 63
 - interactivity, 63
 - motion for geometric figures, 62
 - static graphics, 62
 - static screen shot, 62f
 - take-away message, 61–62
- Correlational psychology, 237
- Correlations claims, 168
- Constraint satisfaction models, 224
- Creature League Task. *See also* Collective search in problem space
 - current solution, 32–33
 - exploration amount, 36
 - group benefits, 37
 - icon's popularity, 35
 - imitators' similarity, 35, 36f
 - mean group score vs. guess proportion, 36–37
 - number of points, 33–34
 - Pokemon-like creatures team, 32, 33f
 - random imitation strategy, 34–35, 34f
 - results and implications, 34
 - scores, 35
- D**
- Deese-Roediger-McDermott (DRM) list-learning paradigm, 78
- Diagrammatic reasoning, 68
- Discovery effects, 122–123
- Domain-general processes, 253
- Domain-specific cognition, 78
- DRM list-learning paradigm. *See* Deese-Roediger-McDermott list-learning paradigm
- Dual-process framework
 - causal illusions, 191–192
 - causal view experiment, 193, 194f
 - causal impression, 194
 - causal judgments, 195–196, 195f
 - results, 196
 - temporal contiguity, 194
 - intuitive process, 192–193
 - potential solution, 193
 - reflective process, 192–193
 - from visual materials, 191
- E**
- Elementary cognitive tasks, 248
- Embodied cognition, 142
 - claims, 143
 - cognitive tasks, 143
 - properties, 143–144
- Embodiment, 97
 - accessing action compatibility, 98–99
 - ACEs, 100–101
 - alternative reference frames, 102–103
 - effects, 100–101
 - mouse tracking task, 103f
 - proximity between locations, 98
 - representational perspective, 102
 - of route and survey perspectives, 104, 105f
 - route perspective mental tour, 106

- Embodiment (*Continued*)
 simulation effects, 107
 southern route bias, 106
 in spatial information retrieval, 99–100
 spatial knowledge development, 99, 100f
 spatial reference frames, 101–102, 104
- Event memory, 95
- Executive attention interpretation, 254f
- Exemplar model, 206–210
 using Bayesian network, 209
 GCM, 208–209
 learning, 208f
- Exploration, 7–8
 and exploitation, 7–9
 individual decision versus group
 performance, 8–9
- Eye-movement heat maps, 85, 86f–87f
- F**
- Factor analysis, 240–241
 development, 241
 latent variables, 241
 Spearman's original model, 242f
- Factor rotation, 242–243
- Feature-based accounts of categorization,
 205–206
 Anderson's model, 206–207, 206f
 exemplar model. *See* Exemplar model
 mixture model, 209–210
 prototype model, 206–208
 using Bayesian network, 208
 learning, 207f
- Fluency effects, 122–124
- Fluid-crystallized model, 246–247
- Flynn-effect, 250
- Force-based accounts. *See also* Causation
 Aristotle's force-based approach, 169–170
 actions of living things, 170
 causal powers theory, 170
 natural mechanism, 171
 Newton's force theory, 172
 as psychological causes, 170
 grounded force-based approaches, 172
 input, 172–173
 kinesthetic sense, 173
 notion of force, 173–174
 personal experiences, 173
 individual causal claims, 169
 problems
 impressions measurement, 197
 participants' rating, 196–197
 rating differences, 197
- Fully connected network structures, 11–12,
 12f
- Fundamental Theorem of Learning, 149
 PL, 148
 relations among types of learning, 148–149
 “sensorimotor” transactions, 148–149
- G**
- Gc. *See* Tests of crystallized intelligence
- GCM. *See* Generalized Context Model
- General factor, 255. *See also* Single general
 factor
 executive attention theory, 255
 general intelligence factor, 248
 interpretation, 256
 multiple domain-general mechanisms,
 255–256
- Generalized Context Model (GCM), 208
- Gestures, 68–69
- Gf. *See* Tests of fluid
- Gf–Gc model, 258
- Global positioning system (GPS), 90
- Goal processing and attention, 84–85
 behavioral goals, 85
 eye-movement heat maps, 85, 86f–87f
 goal-consistent information, 85–87
 goal-relevant information, 87
- GPS. *See* Global positioning system
- Grounded force-based approaches, 172. *See
 also* Aristotle's force-based approach
 input, 172–173
 kinesthetic sense, 173
 notion of force, 171, 173–174
 proprioception (or kinesthesia), 174
 personal experiences, 173
- Group factors. *See also* Single general factor
 general factor and, 244
 psychometric measurement, 243–244
 rotation of factors, 243
 Spearman's model, 242, 242f
 statistical method, 242–243
 Thurstone's model, 243f

- Group performance, 19
 for multimodal function, 20–21, 22f
 for needle function, 21–22, 23f
 for single-peaked function, 20, 21f
- H**
- Higher-dimensional problem spaces,
 collective learning in, 176
 “Draw the Mystery Picture” task,
 176–178
 major results, 178–183
- Higher-order factor models, 244
- Hypothetical processes, 251t
- I**
- In-vehicle technology, 91
- Inattentive blindness, 80
- Inference making
 designers make discoveries, 64–65
 from diagrams and descriptions
 arrows in, 65–66
 car brake, description, 65–66
 effects, 66
 expertise/ability effect, 67–68
 high ability/expertise participants,
 67–68
 performance, 66–67
 structural or functional diagram, 67f
 true/false statements, 66
- Innovation propagation
 agent-based computational model
 group performance, 19
 networks, 22, 24
 Particle Swarm Algorithm, 18–19
 searching problem space, 22–23
 SSEC, 18–19
 conformity, 9–10
 fieldwork, 10–11
 fitness function, 11, 13
 hill-climbing method, 13
 for multimodal fitness functions, 13, 14f,
 15
 needle payout function, 15–17, 16f
 network structures, 11–12, 12f
 participants percentage, 13–15, 15f
 progressions, 17–18, 18f
 for unimodal fitness functions, 13, 14f
- Innovations, 9
- Intelligence, 234
 bi-factor model, 244
 latent variable model, 245f
 content of factors
 bi-factor models, 246
 CHC model, 247
 fluid-crystallized model, 246–247
 radex models, 246
 large-scale analyses, 247
 three-stratum models, 246
- factor analysis, 240–241
 development, 241
 latent variables, 241
- group factors
 general factor and, 244
 psychometric measurement,
 243–244
 rotation of factors, 243
 Spearman’s model, 242, 242f
 statistical method, 242–243
 Thurstone’s model, 243f
- higher-order factor models, 244
 Carroll’s hierarchical latent variable
 model, 244, 245f
- intelligence components, 258–259
 latent variable model, 242f
 psychometric measurement, 243–244
 radex model, 245–246, 246f
 single general factor, 241–242
 evidence against unitary source,
 249–250
 multiple sources, 250–252
 nine hypothetical processes, 251t
 unitary source, 247–249
 and working memory capacity,
 256–257
- Intelligence quotient (IQ), 235
- Interactive learning, 5
- Interference, 255
- IQ. *See* Intelligence quotient
- J**
- Just-in-time models, 83
- K**
- Kinesthesia. *See* Proprioception

L

- Latent variable model
 - Carroll's hierarchical, 245f
 - Spearman's original model, 242f
 - Thurstone's model, 243f
- Learning within learners' community, 3–4
 - bridging between individual and group levels
 - Asch's method, 6
 - dependencies elimination, 6–7
 - individual vs. social learning, 7
 - segregation, 6–7
 - experiments on learning, 2
 - exploitation, 8
 - exploration, 7–8
 - in groups, 4
 - group dynamic patterns, 4
 - parallel problem solving situations, 4–5
 - innovations, 9
 - open source software community, 3
 - social learning, 2–3
- Lego instructions, 57–58
- Life, 47–48. *See also* Perceiving events; Telling events
 - perception, 48
 - spraction, 48
 - abstraction, 72–73
 - action, 72–73
 - space, 72–73
 - stories, 48
- Location/identity binding trials, 81–82

M

- Michotte's arguments
 - causal impression, 179
 - configuration of forces, 180–181
 - forces role in causation perception
 - amplification of motion, 182
 - notion of force, 182–183
 - physical contact effect, 181–182
 - radii of action, 182
 - temporal contiguity, 181–182
 - for independence of launching effect, 179–180
 - object properties, 181

- Mixture model, 209
 - Anderson's model as Bayesian network, 206f, 209–210
 - interpolation between exemplar and prototype models, 209–210
- Modal symbols, 136–137
- Moravec's Paradox, 120
- Mouse tracking task, 103f
- Multimodal fitness functions, 13, 14f
- Murphy and Medin models, 210–211
 - conceptual framework, 211
 - Rehder's model, 212, 226
 - implications of, 213
- Mutualism model, 252

N

- National Assessment of Education Progress, 152–153
- Natural kind learning, 146–148
- Naturalistic events, 50–51
- Navigational aid, 90–91
- Needle payout function, 15–17, 16f
 - network structure performance, 15–17, 17f
- Network structures of learning groups, 11–12, 12f
- Neuroimaging studies, 261
- Newton's force theory, 172
- Nonoverlapping processes, 119

O

- Off-line cognition, 143
- Olympian perceptual abilities, 120
- One-factor latent variable model, 254f
- Open source software community, 3
- Overlapping psychological processes, 251

P

- PALM. *See* Perceptual/adaptive learning module
- Parallel problem solving situations, 4–5
 - influence by others, 6
 - interactive, 5
 - physical constant estimation, 5
 - via public nature, 5
- Parietal operculum, 174

- Particle Swarm Algorithm, 18–19
 single-peaked and multi-peaked fitness functions, 19, 20f
 version, 19
- PEAT. *See* Probability Expectancy Attribution Theory
- Perceiving events, 48–49
 breakpoint, 49
 fundamental characteristic, 49
 incorporate aspects, 51–52
 mean detrended log looking time scores, 51f
 naturalistic events, 50–51
 perfect convergence, 52
 play-by-play descriptions, 49
 segmenting and comprehending events, 52–53
 top-down and bottom-up information, 52
- Perception, 125, 129–130
 classical view, 132
 relationship with cognition, 131
 role in learning mathematics, 131
- Perceptual learning (PL), 119. *See also* Social learning
 amodal and abstract character, 135
 amodal and abstract parallels, 140
 feature activations, 136
 innate foundations, 135
 modal and amodal symbols, 136–137
 orientation-sensitive units, 138–140
 perceptual mechanisms, 137–138
 sensory activations, 136
 amodal perceptual representations, 141
 embodied cognition, 142–144
 evidence and argument, 142
 neurophysiological evidence, 142
 perceptual information, 141
 sensory input channels, 141–142
 auditory discrimination training, 150–151
 character selection, 140
 contemporary views, 141
 information selection in classical views, 140–141
 cognitive tasks, 131
 complex-learning domains, 132
 conventional instruction, 153–154
 effects, 122
 discovery and fluency effects, 122–123
 expert and novice information extraction, 123t
 fluency effects, 123–124
 interpretation, 124–125
 possible signature effects, 124t
 face classifications, 119–120
 growth of perceptual expertise, 150
 implications for, 144
 and instruction, 145
 complex-learning domains, 157–158
 declarative and procedural knowledge, 159
 elements in instruction, 155
 natural kind learning, 146–148
 PL technology, 150–155
 possible moves and countermoves, 145–146
 properties, 156–157
 relations among types of learning, 148–149
 variety of domains, 146
 learning software, 154
 linear measurement PLM, 153f
 in mathematics
 algebraic transformations PLM, 127, 127f
 algebra problem solving, 128, 128f
 arbitrary relation, 129
 broken ruler problem, 129f
 discovery and fluency aspects, 129–130
 distributive property, 126–127
 higher cognitive processes, 125
 integrating methods, 130–131
 perception, 129–130
 and science learning, 151–152
 using single equation, 126
 structures and relations, 125–126
 using tangible representations, 130
 valid transformation, 128
 very high-level cognitive domains, 125
 medical and surgical training applications, 151
 modeling, 150
 Moravec's Paradox, 120

- Perceptual learning (PL) (*Continued*)
- National Assessment of Education Progress, 152–153
 - Olympian perceptual abilities, 120
 - PALMs, 151
 - perception, 132
 - perceptual Olympians, 121
 - problems for understanding
 - Barsalou's PSS view, 134
 - extraction and fluent processing, 135
 - higher-level reasoning, 133
 - massive problems with abstraction and selection, 134–135
 - perception and cognition, 135
 - role in learning and expertise, 119
 - scientific progress and practical applications, 131
 - verbal explanations or procedural calculations, 154
 - in Visual Navigation PLM, 150
 - Perceptual learning module (PLM), 125–126
 - Perceptual Olympians, 121
 - Perceptual symbol systems (PSS), 132–133
 - cognitive processes and perception, 133
 - perceptual encodings, 133
 - re-enactment in modality-specific systems, 133
 - transduced perception, 133
 - Perceptual/adaptive learning module (PALM), 151
 - PL. *See* Perceptual learning
 - Play-by-play descriptions, 49
 - PLM. *See* Perceptual learning module
 - Positive manifold, 241, 248, 252
 - Postcentral gyrus, 174
 - Powers, inactive, 169–170
 - Preattentive feature processing, 81
 - Prediction tests. *See also* Categorization
 - category-irrelevant contextual features, 221t, 223t
 - categorization judgment, 222–223
 - effects on categorization, 220
 - generative causes, 222–224
 - participant rates, 220–222
 - preventative causes, 222–224
 - design, 218–219
 - model predictions, 219–220
 - Present model. *See also* Categorization
 - causal Bayesian network framework, 213–214, 215f
 - category-irrelevant generative causes, 216f
 - external causes, 214
 - features, 217
 - flexibility, 217
 - representation of experimental design, 215f
 - sigmoid function, 216–217
 - variable representing category membership, 214
 - impetus for, 212–213
 - Primary somatosensory cortex (SI), 174
 - Probability Expectancy Attribution Theory (PEAT), 225
 - Proprioception, 174
 - PSS. *See* Perceptual symbol systems
 - Psychology, 234
 - complex span tasks, 236
 - correlational, 237
 - developmental psychologists, 235
 - differential and experimental, 235
 - experimental and correlational, 237
 - experiments on learning, 2
 - interim summary, 240
 - IQ-test, 235
 - Search of PsycInfo, 234
 - skepticism, 236
 - social learning, 2–3
 - unification
 - differential/correlational psychology, 238
 - interindividual and intraindividual lines, 238–239
 - interpretation of latent variables, 238
 - interpretation of psychometric, 238
 - intraindividual processes, 237–238
 - mutual underdetermination, 239–240
 - performance, 239
 - refutable predictions, 240
 - working memory capacity, 236–237
- R**
- Radex models, 245–246, 246f
 - Random network structures, 11–13, 12f
 - Reading span task, 252–253

- Regular lattice network structures, 11–12, 12f
- Representational perspective, 102
- Retinal model, 138–140
- Rotation of factors, 243
- Round-based picture-matching puzzle game, 25–26, 26f
- cell-by-cell comparison, 26–27, 27f
 - diversity, 26–27
 - imitation rate, 30–31
 - imitators' similarity comparison, 29, 30f
 - improvement, 26–27
 - results, 27–28, 28f
 - strategy and performance relationship, 28–29, 29f
 - turnover, 26–27
- S**
- Sampling theory. *See* Thompson's model
- Search of PsycInfo, 234
- Secondary somatosensory cortex (SII), 174
- another person touch activity, 177–178
 - from neural imaging, 176–177
- Self-, Social-, and Exploration-based Choices (SSEC), 18–19
- Sense of force, 183
- in auditory signal control condition, 186
 - causal events, 185–186
 - causal vs. noncausal event, 186–187, 186f
 - correlations, 189–191
 - difference in reaction time, 186–187, 188t
 - dual-process framework
 - causal illusions, 191–192
 - causal view, 193–196
 - intuitive process, 192–193
 - potential solution, 193
 - reflective process, 192–193
 - from visual materials, 191
 - physical force experiments, 183–185, 184f
 - predictions, 190
 - trial structure, 185
 - in visual signal control condition, 186
- Sense of touch, 174
- forces from neural imaging
 - activity in SII, 176–177
 - another person touch activity, 177–178
 - vision and touch, 178
 - forces from psychophysics
 - lifting behavior, 175–176
 - storage of forces, 175–176
 - in tactile system, 174–175
 - forces from weight illusions, 176
 - grounding forces in, 174
- Sensorimotor transactions, 148–149
- Sensory properties, 136–137
- Shape variables, 146–147
- Short-term memory, 260f
- Sigmoidal model, 216–217
- predictions, 219–220
 - requirements, 217
- SII. *See* Secondary somatosensory cortex
- Single general factor, 241–242
- evidence against unitary source, 249–250
 - multiple sources, 250–252
 - nine hypothetical processes, 251t
 - unitary source, 247–249
- Single psychological process, 256
- Skepticism, 236
- Sloped function, positively, 134–135
- Small-world network structures, 11–12, 12f
- Social learning, 2–3
- group-level phenomenon, 40
 - convergence, 40
 - external validity, 42
 - problem space, 41
 - reciprocal imitation, 41
 - imitation heuristics
 - early imitation, 38–39
 - frequency, 38
 - momentum, 39
 - similarity, 39
 - upgrade, 38
 - limitations, 37
- Somatosensory system, 174
- Southern route bias, 106
- Spatial cognition, 78
- attention research, 79
 - arousal and emotion, 89
 - arousal-inducing elements, 88
 - attention distractors, 89–91
 - emotionally arousing information, 88–89

- Spatial cognition (*Continued*)
- goal processing and attention, 84–87
 - implications for automaticity, 80–84
 - levels-of-focus hypothesis, 87–88
 - in situ, 79–80
 - cognition of, 108
 - computer scientists, 54
 - domain-specific cognition, 78
 - memory research, 91
 - event memory, 95
 - spatial memory, 93–96
 - working memory, 91–93
- Spatial information representation, 96–97.
- See also* Amodal perceptual representations
 - ACE, 96–97
 - anecdotal evidence, 97
 - changes as knowledge develops, 97–98
 - embodiment, 97, 107–108
 - accessing action compatibility, 98–99
 - ACEs, 100–101
 - alternative reference frames, 102–103
 - effects, 100–101
 - mouse tracking task, 103f
 - proximity between locations, 98
 - representational perspective, 102
 - of route and survey perspectives, 104, 105f
 - route perspective mental tour, 106
 - simulation effects, 107
 - southern route bias, 106
 - in spatial information retrieval, 99–100
 - spatial knowledge development, 99, 100f
 - spatial reference frames, 101–102
 - summarization, 101
- Spatial memory, 93–96. *See also* Working memory
- category-based errors, 94
 - involving associations, 94
 - organization, 94–95
- Spatial reference frames, 101–102
- Spatial rehearsal effects, 83
- Spearman's model, 242, 242f
- Spraction, 48
 - abstraction, 72–73
 - action, 72–73
 - space, 72–73
- SSEC. *See* Self-, Social-, and Exploration-based Choices
- T**
- Tactile system, 174–175
- Take-away message, 61–62
- Telling events
 - apprehension and congruence principles, 61–63
 - cognitive design principles, 58
 - computer algorithm, 57f
 - gesture for action, 68–69
 - high spatial participants diagram, 54–56
 - incomprehensible instructions, 54
 - inference making
 - designers make discoveries, 64
 - from diagrams and descriptions, 65–68
 - four sketches, 64f
 - instructions
 - to assemble barbeque, 55f
 - to assemble TV cart, 59f
 - maps and depictions and descriptions, 53
 - networks, nodes and links, 53
 - origami and Lego instructions, 54
 - participant assembling, 54–55, 55f
 - participant-produced instructions, 56f
 - proximity still matters, 54
 - real stories, 70
 - static tales, 63–64
 - step-by-step, perspective, action
 - instructions, 58
 - step-by-step perspective action visual
 - instructions, 57–58
 - successes, designing diagrams pair, 60
 - 3-D perspective drawings, 57
 - visual explanations and narratives, 71–72
 - visual story semantics, 60–61
- Tests of crystallized intelligence (Gc), 258
- Tests of fluid (Gf), 258
- Theory-based models, 210
- Thompson's model, 251
- Three-stratum models, 246
- Thurstone's model, 243f

Top-down conceptual information, 50–51

Transfer appropriate processing, 84

Two-factor theory, 241–242

U

Unification

differential/correlational psychology, 238

interindividual and intraindividual lines,
238–239

interpretation of

latent variables, 238

psychometric, 238

intraindividual processes, 237–238

mutual underdetermination, 239–240

performance, 239

refutable predictions, 240

Unified model. *See also* Intelligence

converging evidence, 259–262

intelligence

components, 258–259

and working memory capacity, 256–257

working memory

components, 257–258, 257f

and short-term memory, 260f

Unimodal fitness functions, 13, 14f

Unitary intraindividual process, 249

Unitary source, 247–249

V

Verbal–perceptual–mental–rotation, 247

Visual signal control condition, 186

Visual story semantics, 60–61

Visuospatial sketchpad, 92–93

W

Wechsler Adult Intelligence Scale (WAIS),
260–261

Working memory, 91–92, 234

choreographed role, 93

information integration, 93

subcomponent assessment, 91–92

visuospatial sketchpad role, 92–93

Working memory capacity models

components, 257–258

general factor, 255–256

history repeating

bi-factor model of intelligence, 253

debate and structures, 253

domain-general processes, 253

executive attention interpretation,
254f

one-factor latent variable model,
254f

positive manifold, 252

reading span task, 252–253

This page intentionally left blank