The Psychology of Learning and Motivation

Volume 64

VOLUME SIXTY FOUR

THE PSYCHOLOGY OF LEARNING AND MOTIVATION

Series Editor

BRIAN H. ROSS

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

VOLUME SIXTY FOUR

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
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
125 London Wall, London EC2Y 5AS, UK
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2016

ISBN: 978-0-12-804739-2
ISSN: 0079-7421

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

CONTENTS

This page intentionally left blank

CONTRIBUTORS

Nadia M. Brashier
Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

Guillermo Campitelli
School of Psychology and Social Science, Edith Cowan University, Perth, Australia

Allison D. Cantor
Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

Robert L. Greene
Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, USA

David Z. Hambrick
Department of Psychology, Michigan State University, East Lansing, MI, USA

Eliot Hazeltine
Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA, USA

Todd A. Kahan
Department of Psychology, Bates College, Lewiston, ME, USA

Brooke N. Macnamara
Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, USA

Elizabeth J. Marsh
Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

Daniel Morrow
Department of Educational Psychology, University of Illinois Urbana-Champaign, Champaign, IL, USA

Miriam A. Mosing
Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

Gregory L. Murphy
Department of Psychology, New York University, New York, NY, USA

Diane Pecher
Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands

Eric H. Schumacher
School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

Fredrik Ullén
Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

René Zeelenberg
Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands

This page intentionally left blank

CHAPTER ONE

Beyond Born versus Made: A New Look at Expertise

David Z. Hambrick*,[1], Brooke N. Macnamara[§], Guillermo Campitelli[¶], Fredrik Ullén[||] and Miriam A. Mosing[||]

*Department of Psychology, Michigan State University, East Lansing, MI, USA
[§]Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, USA
[¶]School of Psychology and Social Science, Edith Cowan University, Perth, Australia
[||]Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden
[1]Corresponding author: E-mail: hambric3@msu.edu

Contents

Abstract

Why are some people so much more successful than other people in music, sports, games, business, and other complex domains? This question is the subject of one of psychology's oldest debates. Over 20 years ago, Ericsson, Krampe, and Tesch-Römer (1993) proposed that individual differences in performance in domains such as these largely reflect accumulated amount of "deliberate practice." More controversially, making exceptions only for height and body size, Ericsson et al. explicitly rejected any direct

Psychology of Learning and Motivation, Volume 64
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2015.09.001

role for innate factors ("talent") in the attainment of expert performance. This view has since become the dominant theoretical account of expertise and has filtered into the popular imagination through books such as Malcolm Gladwell's (2008) *Outliers*. Nevertheless, as we discuss in this chapter, evidence from recent research converges on the conclusion that this view is not defensible. Recent meta-analyses have demonstrated that although deliberate practice accounts for a sizeable proportion of the variance in performance in complex domains, it consistently leaves an even larger proportion of the variance unexplained and potentially explainable by other factors. In light of this evidence, we offer a "new look" at expertise that takes into account a wide range of factors.

1. INTRODUCTION

No one can deny that some people are vastly more skilled than other people in certain domains. Consider that the winning time for the New York City Marathon in 2014—just under 2 h and 11 min—was more than *2 h* better than the average finishing time (http://www.tcsnycmarathon.org/results). Or consider that Jonas von Essen, en route to winning the 2014 World Memory Championships, memorized 26 decks of cards *in an hour* (http://www.world-memory-statistics.com).

What are the origins of this striking variability in human expertise?[1] Why are some people so much better at certain tasks than other people? One particularly influential theoretical account attempts to explain individual differences in expertise in terms of *deliberate practice* (e.g., Boot & Ericsson, 2013; Ericsson, 2007; Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson, Nandagopal, & Roring, 2005; Keith & Ericsson, 2007). Here, we describe the mounting evidence that challenges this view. This evidence converges on the conclusion that deliberate practice is an important piece of the expertise puzzle, but not the only piece, or even necessarily the largest piece. In light of this evidence, we offer a "new look" at expertise that takes into account a wide range of factors, including those known to be substantially heritable.

The rest of the chapter is organized into the following sections. We describe the deliberate practice view (Section 2) and then review evidence that challenges it (Section 3). Then, we review evidence for factors other than deliberate practice that may also account for individual differences in

[1] Throughout this chapter, we use the term *expertise* to refer to performance within a particular domain (i.e., domain-specific performance).

expertise (Section 4). We then describe an integrative approach to research on expertise (Section 5). Finally, we summarize our major findings and comment on directions for future research (Section 6).

2. THE DELIBERATE PRACTICE VIEW

The question of what explains individual differences in expertise is the topic of one of psychology's oldest debates. One view is that experts are "born." This view holds that although training is necessary to become an expert, innate ability—*talent*—limits the ultimate level of performance that a person can achieve in a domain. Nearly 150 years ago, in his book *Hereditary Genius,* Francis Galton (1869) argued for this view based on his finding that eminence in domains such as music, science, literature, and art tends to run in families, going so far as to conclude that "social hindrances cannot impede men of high ability, from becoming eminent [and] social advantages are incompetent to give that status, to a man of moderate ability" (p. 41). The opposing view is that experts are "made." This view argues that if talent exists at all, its effects are overshadowed by training. John Watson (1930), the founder of behaviorism, championed this view when he guaranteed that he could take any infant at random and train him to become "any type of specialist [he] might select…regardless of his talents" (p. 104).

The modern era of scientific research on expertise traces back to the 1940s and the research of the Dutch psychologist Adriaan de Groot (1946/1978). Himself an internationally competitive chess player, de Groot investigated the thought processes underlying chess expertise using a "choice-of-move" paradigm in which he gave chess players chess positions and instructed them to verbalize their thoughts as they considered what move to make. From analyses of their verbal reports, de Groot discovered that there was no association between skill level and the number of moves ahead a player thought in advance of the current move. Instead, he found evidence for a perceptual basis of chess expertise. As de Groot put it, the grandmaster "immediately 'sees' the core of the problem in the position" whereas the weaker player "finds it with difficulty—or misses it completely" (p. 320). de Groot attributed this ability to a "connoisseurship" (p. 321) that develops through years of experience playing the game.

Nearly 30 years later, de Groot's (1946/1978) work was the inspiration for Chase and Simon's (1973a) classic study of chess expertise, which marks the beginning of cognitive psychologists' interest in expertise. Testing three chess players—a master, an intermediate-level player, and a beginner—Chase and Simon found that there was a positive relationship between chess skill and memory for chess positions, but only when they were plausible game positions. When the positions were random arrangements of pieces, there was almost no effect of chess skill on memory. Based on these findings, Chase and Simon (1973b) concluded that although "there clearly must be a set of specific aptitudes...that together comprise a talent for chess, individual differences in such aptitudes are largely overshadowed by immense individual differences in chess experience. Hence, the overriding factor in chess skill is practice" (p. 279).

The experts-are-made view has held sway in the scientific literature ever since. Over 20 years ago, in a pivotal article, Ericsson et al. (1993) proposed that individual differences in performance in complex domains (music, chess, sports, etc.) largely reflect differences in the amount of time people have spent engaging in *deliberate practice,* which "includes activities that have been specially designed to improve the current level of performance" (p. 368). In the first of two studies, Ericsson et al. recruited violinists from a Berlin music academy and asked them to estimate the amount of hours per week they had devoted to deliberate practice since taking up the violin. The "best" violinists had accumulated an average of over 10,000 h of deliberate practice by age 20, which was about 2500 h more than the average for the "good" violinists and about 5000 h more than the average for the least accomplished "teacher" group. In a second study, Ericsson et al. found that "expert" pianists, who were selected to be similar in skill level to the good violinists in the first study, had accumulated an average of over 10,000 h of deliberate practice by age 20, compared to only about 2000 h for "amateur" pianists (see Ericsson, 2006; for further discussion of these results).

Ericsson et al. (1993) concluded that "high levels of deliberate practice are necessary to attain expert level performance" (p. 392). More controversially, they added:

> *Our theoretical framework can also provide a* sufficient account *of the major facts about the nature and scarcity of exceptional performance. Our account does not depend on scarcity of innate ability (talent) and hence agrees better with the earlier reviewed findings of poor predictability of final performance by ability tests. We attribute the dramatic differences in performance between experts and amateurs-novices to similarly large differences in the recorded amounts of deliberate practice.*
>
> ***Ericsson et al., (1993, p. 392), emphasis added***

Ericsson et al. further claimed that "individual differences in ultimate performance can largely be accounted for by differential amounts of past and current levels of practice" (p. 392), and stated:

> *We agree that expert performance is qualitatively different from normal performance and even that expert performers have characteristics and abilities that are qualitatively different from or at least outside the range of those of normal adults. However, we deny that these differences are immutable, that is, due to innate talent. Only a few exceptions, most notably height, are genetically prescribed. Instead, we argue that the differences between expert performers and normal adults reflect a life-long period of deliberate effort to improve performance in a specific domain.*
> *(p. 400)*

Ericsson and colleagues have maintained their view over the past two decades. Ericsson et al. (2005) explained:

> *the individual differences in genetically determined capacities and fixed structures required for the development of elite performance appear to be quite limited, perhaps even restricted, to a small number of physical characteristics, such as height and body size. The expert performance framework attempts to explain the large individual differences in performance in terms of individual differences in sustained deliberate practice.*
> *(p. 305)*

Similarly, Keith and Ericsson (2007) argued that "an individual's level of performance in a particular domain is the result of effortful practice activities in which he or she has engaged in over the course of several years with the explicit goal of performance improvement" (p. 135), and clarified that deliberate practice "activities can be designed by external agents, such as teachers or trainers, or by the performers themselves" (p. 136; see also Ericsson, 1998, for this point). Ericsson (2007) claimed that "it is possible to account for the development of elite performance among healthy children without recourse to unique talent (genetic endowment)—excepting the innate determinants of body size" (p. 4), and reflected: "My own thoughts on exceptional ability were influenced by my family and education in Sweden, where views that genetic endowment limited the acquisition of superior performance among otherwise healthy individuals were discouraged." (p. 5).

3. CHALLENGES TO THE DELIBERATE PRACTICE VIEW

It is difficult to overstate the impact of the deliberate practice view. At the time of this writing, the Ericsson et al. (1993) article has been cited over 5400 times (Source: Google Scholar), making it one of the most cited articles in the psychological literature, and nearly a hundred theses and dissertations have been conducted on deliberate practice over the past two decades (Source: ProQuest Dissertations & Theses Global). Citing Ericsson and colleagues' research, one of us noted in a *New York Times* op-ed that there is no denying the "power of practice" (Hambrick & Meinz, 2011a).

Ericsson and colleagues' findings have also filtered into popular culture. Most notably, Ericsson et al.′s (1993) findings were the inspiration for what the writer Malcolm Gladwell termed the "10,000 hour rule" in his bestselling book *Outliers* (2008)—the idea that it takes 10,000 h to become an expert. The 10,000 h rule has since inspired thousands of internet articles and blog posts, and even a rap song that was the theme music for a *Dr Pepper* commercial.[2] No psychologist has had a greater impact on the public's view of expertise than Ericsson.

Nonetheless, it seems fair to say that Ericsson and colleagues' view has been met with considerable skepticism in the scientific literature. Gardner (1995) commented that Ericsson and colleagues' view "requires a blindness to ordinary experience—as well as to decades of psychological theorizing" (p. 802; for a reply, see Ericsson & Charness, 1995), and Schneider (1998) noted that he was "very sympathetic to the model of skill acquisition initially developed by Ericsson and colleagues" but questioned the "basic assumption that progress in a given domain is solely a function of deliberate practice" (p. 424). Winner (2000) observed that "Ericsson's research demonstrated the importance of hard work but did not rule out the role of innate ability" (p. 160), and Anderson (2000) stated that "Ericsson and Krampe's research does not really establish the case that a great deal of practice is sufficient for great talent" (p. 324). Detterman, Gabriel, and Ruthsatz (1998) described the position advocated by Ericsson and colleagues as "absurd environmentalism" (p. 411).

More recently, Gagné (2007, 2013) criticized Ericsson for misrepresenting evidence contrary to his (Ericsson's) view and for caricaturing opposing positions so as to create "straw men" (for a reply, see Ericsson, 2013a), and Tucker and Collins (2012) noted that Ericsson "overlooks a body of

[2] Ericsson has discussed the 10-year rule extensively (e.g., Ericsson et al., 1993; Boot & Ericsson, 2013), but has emphasized that the 10,000-*hour* rule was invented by Malcolm Gladwell, and that the findings from his (Ericsson's) research were only the "stimulus" for the 10,000-hour rule (see Ericsson, 2012). We do not attribute the 10,000-hour rule to Ericsson. For comment by Ericsson on the 10,000-hour rule, see: https://web.archive.org/web/20150614160055/http://www.abc.net.au/radionational/programs/allinthemind/practice-makes-perfect/3611212#.

scientific literature which strongly disproves his model" (p. 555; for a reply, see Ericsson, 2013b). Marcus (2012) wrote:

> *The psychologist Anders Ericsson went so far as to write, 'New research shows that outstanding performance is the product of years of deliberate practice and coaching, not of any innate talent or skill.' How I wish it were true…. Practice does indeed matter—a lot—and in surprising ways. But it would be a logical error to infer from the importance of practice that talent is somehow irrelevant, as if the two were in mutual opposition.*
> *(p. 97)*

Ackerman (2014) added that "until Ericsson shows cognitive expertise development in a randomly selected group of subjects, including those with moderate mental retardation, there is no reason to believe that such development can be accomplished" (p. 105).

Other scientists have criticized Ericsson and colleagues' methodological approach—the *expert performance approach* (see Boot & Ericsson, 2013; Ericsson & Smith, 1991). Noting that reputation, credentials, and years of experience may correlate weakly with actual performance in a domain, Ericsson and colleagues have emphasized the importance of measuring expertise under controlled conditions using laboratory tasks representative of a domain. The paradigmatic example is the choice-of-move task from de Groot's (1946/1978) chess research. However, Hoffman et al. (2014) have argued that restriction of expertise research to laboratory tasks removes many important professions from consideration, including those in which it is not possible or practical to devise laboratory tasks to capture the essence of expertise in the domain (e.g., astronaut; see also Weiss & Shanteau, 2014). More generally, Wai (2014) noted that "Ericsson appears unable to go beyond his own framework and definitions to incorporate the approaches of others as well as the full network of evidence surrounding the development of expertise" (p. 122).

Thus, although Ericsson and colleagues' view has had enormous impact on both scientific and popular views of expertise, it has been sharply criticized on both conceptual and methodological grounds in the scientific literature.

3.1 Empirically Evaluating the Deliberate Practice View

We have challenged the deliberate practice view on empirical grounds. The major question we have tried to address in our research is simply how important deliberate practice is as a predictor of individual differences in expertise. That is, can individual differences in domain-specific performance largely be accounted for by accumulated amount of deliberate practice, as Ericsson and colleagues have argued?

To answer this question, Hambrick, Oswald, et al. (2014) performed a reanalysis of studies of music and chess, two of the most popular domains for research on expertise. There were two criteria for including a study in the reanalysis: (1) continuous measures of some activity interpretable as deliberate practice and of domain-specific performance were collected, and (2) a correlation between the measures was reported. Hambrick et al. identified six studies of chess and eight studies of music that met these criteria. Ericsson (2013b) noted that correlations between deliberate practice and performance underestimate the true relationship between the two variables, because neither variable can be assumed to be perfectly reliable:

> *The collected reliability of cumulated life-time practice at different test occasions in large samples has typically been found to range between 0.7 and 0.8 implying that estimates of training history could never account for more than 49–64% of variance in measures of performance—even less for measures of performance that are not perfectly reliable.*
> *(p. 534)*[3]

Therefore, using the standard psychometric approach (Hunter & Schmidt, 1990), Hambrick et al. corrected each correlation for the unreliability of both deliberate practice and performance, and asked specifically how much of the *reliable* variance in performance does deliberate practice explain.

Not surprisingly, deliberate practice and performance correlated positively in all of the studies included in the reanalysis. However, even after correcting for unreliability, the correlations indicated that deliberate practice left more of the variance in performance unexplained than it explained. To be exact, as shown in Figure 1, the average proportion of reliable variance in performance explained was 34% for chess and 29.9% for music. Thus, deliberate practice did not largely account for individual differences in expertise in either domain. In a subsequent meta-analysis of a larger number of music

[3] Ericsson's (2013b) point that less-than-perfect reliability attenuates correlations is correct. However, per the standard formula for a correlation in classical measurement theory ($r_{xy} = r_{x_t y_t}(r_{xx} r_{yy})^{1/2}$, where r_{xy} is the observed correlation, $r_{x_t y_t}$ is the correlation between the "true" scores, and r_{xx} and r_{yy} are the reliabilities of x and y, respectively; see Schmidt & Hunter, 1999), if the reliability of one variable (e.g., deliberate practice) ranges from 0.70 to 0.80, then it could never be expected to account for more than 70–80% of the variance in the other variable (e.g., performance), not 49–64%, and even less if the other variable is not perfectly reliable.

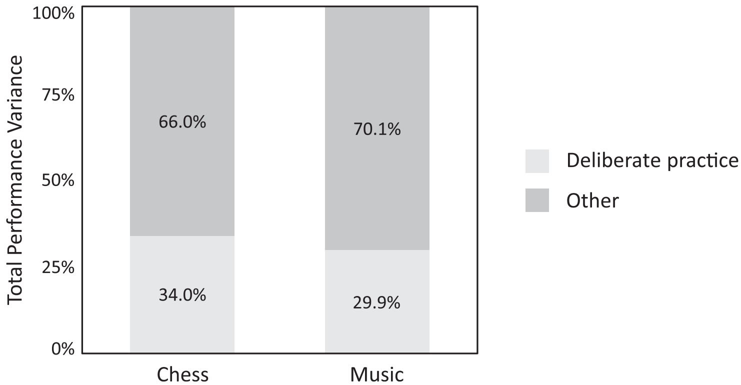

Figure 1 Average percentage of variance in chess performance (left) and music performance (right) accounted for by deliberate practice, correcting for measurement error. The light gray region represents reliable variance explained by deliberate practice; the dark gray region represents reliable variance not explained by deliberate practice. *Adapted with permission of Elsevier from Hambrick, Oswald, et al. (2014), Figures 1 and 3.*

studies, Platz, Kopiez, Lehmann, and Wolf (2014) found that deliberate practice explained 36% of the reliable variance in music performance (avg. corrected $r = 0.61$).

In a commentary, Ericsson (2014a) claimed that Hambrick, Oswald, et al. (2014) rejected his view based on a "common sense basis" (p. 98). In a published reply, Hambrick, Altmann, et al. (2014) explained that they rejected the deliberate practice view on an *empirical* basis—the finding that deliberate practice does not largely account for individual differences in expertise in two of the most widely studied domains in research on expertise. Ericsson also criticized Hambrick, Oswald et al.'s (2014) analysis for ignoring "the effects of forgetting, injuries, and accidents, along with the differential effects of different types of practice at different ages and levels of expert performance" (p. 84). Hambrick, Altmann, et al. (2014) pointed out that Ericsson has never considered all of these factors in his own studies and that their reanalysis included studies that Ericsson has explicitly praised and used to argue for the importance of deliberate practice (e.g., Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005).

Macnamara, Hambrick, and Oswald (2014) have since performed a meta-analysis that covers all of the major domains in which the relationship between deliberate practice and expertise has been studied: games, music, sports, education, and professions. To be included in the meta-analysis, a study had to collect measures of one or more activities interpretable as reflecting deliberate practice (i.e., an activity specifically created to improve performance in a domain) and refer to at least one publication on deliberate practice by Ericsson and colleagues to place the study in the deliberate practice literature. A study also had to collect a measure of performance reflecting skill in a particular domain and report an effect size reflecting the relationship between that measure and deliberate practice (or provide information necessary to compute an effect size).[4] Macnamara et al. allowed that deliberate practice could be either self-directed or teacher-directed, consistent with Keith and Ericsson's (2007) aforementioned point that deliberate practice activities can be designed by external agents or by performers themselves, and with how Ericsson and colleagues have operationally defined deliberate practice in their own research (as discussed in more detail below).

Through a search of over 9300 documents, Macnamara et al. (2014) identified 88 studies that met these criteria, with a total of 157 effect sizes, and a total sample size of over 11,000. Nearly all of these effect sizes were positive, indicating that high levels of deliberate practice are associated with high levels of performance. But, again, the results indicated that deliberate practice left more of the variance in performance unexplained than it explained. To be exact, on average, deliberate practice explained 12% of the variance, leaving 88% unexplained. Macnamara et al. did not correct individual effect sizes for unreliability, because very few studies in the meta-analysis reported a reliability estimate for both deliberate practice and performance. However, they did correct average effect sizes from the meta-analysis, and across a wide range of reliability assumptions, deliberate practice still explained well less than half of the variance in performance.

Moderator analyses revealed that the effect of deliberate practice was strongest for games (26%), music (21%), and sports (18%), and much weaker for education (4%) and professions (<1%, and not statistically significant). The effect sizes for education and professions may be smaller because deliberate practice is less well defined in these domains and/or because the participants in these studies differed in the amount of prestudy expertise, and thus in the amount of deliberate practice necessary to reach a given level of skill. The relationship between deliberate practice and performance also

[4] The data file for Macnamara et al. (2014) is openly available at https://osf.io/rhfsk.

tended to be larger for activities in which the task environment is highly predictable (e.g., running) than for activities in which the task environment is less predictable (e.g., handling an aviation emergency). This finding is consistent with laboratory research showing that training has a greater impact on performance in predictable tasks than less predictable tasks (e.g., consistently- vs variably-mapped tasks; see Ackerman, 1987).

Moderator analyses further revealed that studies that relied on retrospective estimates of deliberate practice reported higher effect sizes than studies that used a log method in which activity was recorded on an ongoing basis. Indeed, deliberate practice explained 20% of the variance in performance for studies that used a retrospective interview, compared to 12% for studies that used a retrospective questionnaire and only 5% for those that used a log method. This finding suggests that the relationship between deliberate practice and performance may be weaker than what our meta-analysis indicates. That is, the log method presumably yields more valid estimates of deliberate practice than retrospective methods, given that people do not have perfect memory for the past. Ericsson alluded to this point about validity as follows:

> *With better research using daily practice diaries during the entire development of music and chess performance, we might find that individual differences in the amount and timing of deliberate practice [do] not account for all observed variance, but current data cannot claim to show that.*
>
> **(as quoted in Szalavitz, 2013[5])**

Finally, considering the type of performance measure, the relationship between deliberate practice and performance was considerably weaker for studies that used an objective measure of performance—either a standardized measure (e.g., chess rating; avg. $r = 0.28$) or a laboratory task (avg. $r = 0.37$)—than for studies that used group membership (avg. $r = 0.51$). If using an objective measure of performance is ideal for expertise research, this finding further suggests that the true relationship between deliberate practice and performance is weaker than has often been claimed.

[5] This quotation is from a popular article (see http://web.archive.org/web/20150731145946/http://healthland.time.com/2013/05/20/10000-hours-may-not-make-a-master-after-all/). Because quotations in popular articles are sometimes not verbatim and may misrepresent the views of the person quoted, we e-mailed the journalist who wrote the article (Maia Szalavitz) to verify the accuracy of this quotation. She confirmed that the quotation is verbatim from an e-mail she received from K. Anders Ericsson, except the word in brackets (Maia Szalavitz, personal communication, June 4, 2013).

In an even more recent meta-analysis, Macnamara, Moreau, and Hambrick (2015) found that the relationship between deliberate practice and sports performance varied by skill level. Specifically, deliberate practice explained only 1% of the variance in performance for studies that used elite-level athletes (e.g., Olympians vs national-level performers), compared to 19% for studies that used sub-elite athletes, and 29% for studies that used mixed samples with both elite and sub-elite athletes. This finding is inconsistent with the claim that "[i]ndividual differences, even among elite performers, are closely related to assessed amounts of deliberate practice" (Ericsson et al., 1993, p. 363), and instead suggests that deliberate practice may lose its predictive power at elite levels of performance.

Ericsson (2014b) has dismissed the results of Macnamara et al.'s (2014) meta-analysis, arguing that only *one* of the 88 studies (or 1 out of 157 effect sizes) that was included meets his criteria for accurately estimating the relationship between accumulated deliberate practice and performance (see also Ericsson, 2014c; for the supplemental material for this commentary). The one study he accepts is Ericsson et al.'s (1993) second study (the study of pianists). However, Ericsson again rejects studies that he has explicitly cited as support for the importance of deliberate practice in the past, including some of his own studies. For example, he rejects his study of darts (Duffy, Baluch, & Ericsson, 2004) because there was no record of a teacher or coach supervising and guiding all or most of the practice. Yet, he and his colleagues explicitly and repeatedly referred to measures that they collected in this study as measures of "deliberate practice" (see, e.g., Duffy et al.'s (2004) Table 3, p. 240) and concluded that the finding of large differences between expert and novice dart players in these measures "supports one of the main tenets of Ericsson et al.'s (1993) theory whereby expertise is acquired through a vast number of hours spent engaging in activities purely designed to improve performance, i.e., deliberate practice" (p. 243).[6]

Ericsson (2014b) rejects studies by other researchers that he has used to support the deliberate practice view in the past, as well. For example, he rejects Charness et al.'s (2005) study of chess, again because there was no record of a teacher. Yet, he once stated that this study "reports the most

[6] Not even in the report of Ericsson et al.'s (1993) study of pianists, or in the biographical interview that was used in this study (see Krampe, 1994; Appendix A, "Retrospective Estimates for Past Amounts of Practice Alone"), can we find any record that the participants were asked to restrict their practice estimates to *only* activities that were supervised and guided by a teacher.

compelling and detailed evidence for how designed training (deliberate practice) is the crucial factor in developing expert chess performance" (Ericsson, 2005, p. 237). For the same reason, he rejects Sonnentag and Kleine's (2000) study of insurance agents, even though he once explained that "[i]n a study of insurance agents Sonnentag and Kleinc [sic] (2000) found that engagement in deliberate practice predicted higher performance ratings" (Ericsson, 2006, p. 695). We credit Ericsson for his vigorous defense of his view, but we do not believe it is acceptable to use studies to argue for the importance of deliberate practice, and then later reject those studies on the grounds that they did not actually measure deliberate practice.

Ericsson (2014b) makes two more general points in his commentary that bear on the deliberate practice view. First, he states:

> *I have never claimed that deliberate practice can explain all reliable variance in attained performance….On the contrary I have acknowledged for decades that height and body size…cannot be changed by training, yet influence the attainment of elite performance in some domains of expertise.*
>
> ***(Ericsson, 2014b, pp. 5–6)***

However, even in domains in which it is not reasonable to argue that height and body size are factors in performance, the available evidence indicates that deliberate practice leaves a large amount of the variance in expertise unexplained. The most obvious example of such a domain is chess. In Charness et al.'s (2005) aforementioned studies of chess, the higher of the two correlations between deliberate practice and performance in these studies was 0.54 before correction for unreliability and 0.63 after correction (see Hambrick, Oswald, et al.'s, 2014, Table 1). Thus, deliberate practice explained about 40% of the reliable variance in chess rating in that study (i.e., $0.63^2 \times 100 = 39.7\%$), leaving 60% unexplained.

Second, Ericsson (2014b) argues that the correlation between estimated amount and *actual* amount of deliberate practice may range from 0 to nearly 1.0—in other words, that estimates of deliberate practice are "contaminated" to some unknown degree by activities not meeting the criteria for deliberate practice. He explains:

> *The duration of deliberate practice may be correlated with the total duration of practice alone with a correlation ranging from 0.0 to almost 1.0 depending on age and skill level of performer and the particular domain of expertise. However, until studies have successfully measured these correlations it is not possible to estimate the proportion of deliberate practice from estimates of practice alone.*
>
> ***(Ericsson, 2014b, p. 5)***

However, the measure of deliberate practice in the one study that Ericsson argues can be used to accurately estimate the relationship between deliberate practice and performance—Ericsson et al.'s (1993) study of pianists—was total duration of practice alone. If it is not yet known what proportion of this measure is *actual* deliberate practice, as opposed to other activities, then all that can be concluded based on the results of that study (or any other study to date) is that deliberate practice accounts for somewhere between 0% and 100% of the variance in performance—and thus that there is no scientific evidence at all that deliberate practice accounts for individual differences in expertise. Even if the measure of deliberate practice in Ericsson et al.'s study of pianists was in some non-obvious way "purer" than measures of deliberate practice in all of the other studies that have been conducted since, this would mean that the case for the importance of deliberate practice rests largely, or entirely, on the results of a single study with a total sample size of only 24.

Our take is that deliberate practice—as it has been operationally defined and measured in research over the past two decades by Ericsson and colleagues and by others who have used their research as a model—explains a sizeable amount of the variance in expertise, but leaves an even larger amount unexplained. Thus, while the deliberate practice view offers a parsimonious account of expertise, it is not supported by the available empirical evidence. To be sure, crucial questions about the relationship between deliberate practice and performance remain, such as why the relationship appears to be stronger for studies that use a retrospective method to measure deliberate practice than for those that use a log method. One possible explanation for this finding is that when asked to retrospectively estimate deliberate practice, people rely on current level of skill rather than on accurate recollections of past engagement in practice. This could lead to inflated estimates of the relationship between deliberate practice and expertise. Nevertheless, we think it is unlikely that the true relationship between deliberate practice and performance will ultimately be found to be zero or trivially small.

3.2 Findings from Individual Studies

The results of individual studies are consistent with this conclusion. In their exemplary studies, Charness et al. (2005) had chess players provide estimates of serious chess activity and calculated measures of both the accumulated amount of these activities as well as amount in the most recent year. In addition, participants reported the number of years of private chess instruction and number of years of group lessons. For each study, and for a combined data set ($N = 375$), Charness et al. regressed chess rating onto these variables. Variance in chess rating accounted for was 41% for the first study, 31% for the second study, and 34% in the combined data set. In a study of 90 chess players, Gobet and Campitelli (2007) found a weaker, but still significant and sizeable, positive relationship between individual deliberate practice and chess rating ($r = 0.42$, or 17.6% of the variance). Moreover, there was a large amount of variability in deliberate practice, even among the most highly skilled players in the sample. Indeed, one player became a chess master after just over 728 h of individual deliberate practice, while it took another player over 16,000 h (see Campitelli & Gobet, 2011, for further discussion). For total deliberate practice, which included individual and group practice, the range was from 3016 to 23,608 h ($r = 0.57$ with chess rating).

In another impressive study, Howard (2012) collected estimates of engagement in chess-related activities from 533 chess players, ranging in skill from intermediate to grandmaster. Howard found that, along with starting age, a set of practice and other experiential variables accounted for 49% of the variance in chess rating. Total number of tournament games (log) was the strongest single predictor of chess rating ($r = 0.62$; $r = 0.33$ for log total study hours). One potential problem with Howard's study is that he used an internet survey instead of in-person experience interviews (see Ericsson & Moxley, 2012). However, averages for the experience variables were very similar to those obtained through in-person interviews in Charness et al.'s (2005) studies. It could also be argued that in-person interviews introduce experimenter bias that internet surveys do not, and thus that the latter approach is superior for collecting at least certain types of information.

The preceding studies used a cross-sectional design in which participants differing in expertise were tested within a narrow band of time. The obvious advantage of this design over a longitudinal design is that it allows researchers to investigate individual differences in expertise without having to wait months, years, or even decades for the participants to reach their final level of skill. Nevertheless, as Sternberg (1996) reminded, correlation does not imply causation: "deliberate practice may be correlated with success because it is a proxy for ability: We stop doing what we do not do well and feel unrewarded for" (p. 350). Similarly, commenting on Ericsson and colleagues' finding of a correlation between deliberate practice and skill level in music, Winner (2000) observed, "Hard work and innate ability have not been unconfounded" (p. 160).

de Bruin, Smits, Rikers, and Schmidt (2008) investigated this issue by performing a longitudinal analysis comparing Dutch chess players who were enrolled in a national chess training program, but dropped out ("drop-outs"), to players who had remained in the program ("persisters"). There was no difference in the effect of deliberate practice on chess rating in the two groups, leading de Bruin et al. to conclude that "those who ultimately arrive at expert level in chess do so not because of a predisposition to perform deliberate practice more efficiently, but because they put in more hours of deliberate practice" (p. 494). Based on this evidence, Ericsson and Towne (2010) argued against the hypothesis that the correlation between deliberate practice and chess expertise is an artifact of drop-outs. However, it is critical to note that the "drop-outs" in this study had only dropped out of a training program for elite chess players. de Bruin et al.'s analysis does not speak to the critical question of whether people quit chess much earlier (e.g., after 50–100 h of training) because of lack of ability. Thus, Sternberg's (1996) and Winner's (2000) point that correlations between deliberate practice and expertise may be inflated due to selective drop-out remains an important caveat to conclusions about the importance of deliberate practice based on cross-sectional findings.

Two recent case studies of chess further challenge the primacy of deliberate practice. Howard (2011) used biographical and autobiographical sources, along with publicly available chess ratings, to investigate the link between practice and chess skill in the Polgár sisters. Starting at a young age, under the supervision of their father, Susan, Sofia, and Judit Polgár received intensive chess instruction on a near-daily basis. Howard found that the sisters differed both in the highest rating they achieved and in the amount of practice they accumulated to reach that rating. For example, one of the sisters reached a rating of 2735 in an estimated 59,904 h of practice, whereas another peaked at 2577—more than a standard deviation lower—in an estimated 79,248 h of practice. Howard also found that the two sisters who became grandmasters had accumulated a great deal more practice by the time they reached their peak rating than had the eight grandmasters in his sample who reached top-ten in the world ($M = 14{,}021$ h, $SD = 7374$ h). In the other case study, Gobet and Ereku (2014) examined the success of Magnus Carlsen—the highest rated chess player in the world by a wide margin—and found that he had significantly *fewer,* not more, years of deliberate practice than the next 10 best players in the world, even using a starting age that is conservative by three years (age 5, when Carlsen learned the moves, instead of age 8, when he has noted he started playing the game seriously).

SCRABBLE has also been used in a few studies of expertise. Using official SCRABBLE ratings as an index of skill, Tuffiash, Roring, and Ericsson (2007) recruited samples of "elite" and "average" SCRABBLE players and had them provide estimates of engagement in various SCRABBLE-related activities, including an activity that would seem to meet the theoretical description of deliberate practice—serious study. (The elite players were representative of players in the top division of the National SCRABBLE Championship, whereas the average players were representative of the average player in the National SCRABBLE Association.) Although the elite group had accumulated more serious study than the average group, for both groups, the standard deviations for serious study were very similar to the means: average group ($M = 1318$, $SD = 1465$) and elite group ($M = 5084$, $SD = 4818$). This indicates that there was a large amount of variability in the data. As for chess, it appears that people differ greatly in the amount of deliberate practice they require to reach a given level of skill in SCRABBLE.

Research on music further challenges the deliberate practice view. In a study by Sloboda and colleagues (see Sloboda, 1996; Sloboda, Davidson, Howe, & Moore, 1996) that Ericsson has cited to support the importance of deliberate practice, students at a selective music school ("high achievers") were found to have accumulated more "formal practice" than students who were learning an instrument at a nonmusic school ("players for pleasure"). However, Sloboda et al. (1996) noted that there were some students at each skill level who did "less than 20% of the mean amount of practice" and others who did "over four times as much practice than average" (p. 301), and added "it appears that there are a few individuals in all groups who manage to attain grade examination passes on very little practice" (p. 301).

Moreover, in Ericsson et al.'s (1993) study of pianists, accumulated deliberate practice ranged from about 10,000 to over 30,000 h among the expert group (see Figure 2). The expert pianists ranged in age from 20 to 31, and thus some of this variability in deliberate practice was presumably due to age (i.e., more deliberate practice for the older pianists). However, the most practiced expert could have been no more than 11 years older than the least practiced expert, and yet the difference in deliberate practice between these subjects was about 20,000 h. At 4 h a day, a person would have to practice nearly 14 years without missing a single day to accumulate this amount of deliberate practice. Thus, it seems likely that some of the pianists in Ericsson et al.'s sample required much less deliberate practice than others to become experts. Ericsson et al. did report extremely high correlations between

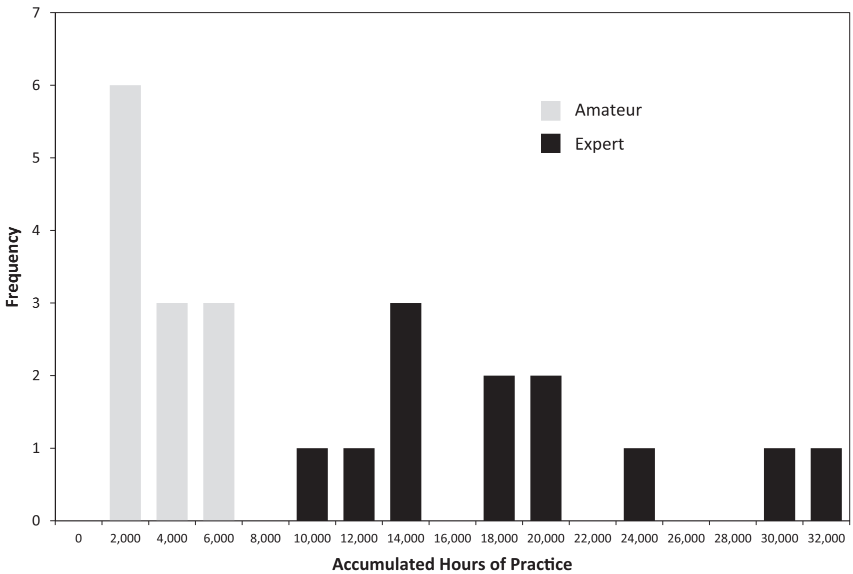

Figure 2 Histogram showing range of deliberate practice for amateur pianists (light gray bars) and expert pianists (dark gray bars) in Ericsson et al. (1993, Study 2). The values used to generate this histogram come from a scatterplot in Ericsson et al.'s Figure 15 (right panel). The first author of this chapter (Hambrick) requested data from the authors of the study, but they were unable to provide it because it is stored on magnetic tape for mainframe computers (Ralf Krampe, personal communication, December 5, 2011). Thus, we extracted the log values from Ericsson et al.'s Figure 15 using Dagra's graphical extraction software (Version 2.0), and then reversed the values to hours (i.e., hours of deliberate practice $= 10^{\text{Log hours}}$). The correlation between the extracted log values and the performance values matches the correlation in Ericsson et al.'s Figure 15 (right panel) exactly ($r = -0.857$). Means are not reported for this variable in Ericsson et al., but the means for the extracted values are very similar to those found in other reports of this study (Krampe, 1994; Krampe & Ericsson, 1996). Thus, we assume that the extracted values accurately capture the variability in the data. In Ericsson et al.'s Figure 15, the variable is labeled "Log-accumulated practice (hours)". We assume that this variable can be interpreted as *deliberate* practice, because elsewhere Ericsson and colleagues describe it as such (see Law, Côté, & Ericsson, 2007).

deliberate practice and performance in a piano-related task ($rs > |0.85|$). However, it must be assumed that these correlations are highly inflated, because an extreme-groups design was used in this study (see Preacher, Rucker, MacCallum, & Nicewander, 2005; for a discussion of issues with extreme-groups designs).

There has also been an extensive amount of research on expertise in sports. Johnson, Tenenbaum, and Edmonds (2006) compared the training histories of elite and sub-elite swimmers. Five of the elite swimmers had won at least one Olympic gold medal, and the other three had been ranked in the top five in the world. The sub-elite swimmers did not meet these lofty criteria, but were still highly accomplished, having participated in national events such as the NCAA championship. Not surprisingly, all of the swimmers had accumulated a large amount of deliberate practice. The overall average was about 7500 h. However, the difference between the groups was not significantly different. In fact, if anything, the mean was higher for the sub-elites (7819 h) than for the elites (7129 h). Furthermore, there was a large amount of variability in amount of deliberate practice. One of the elites—winner of Olympic gold in 1996 and 2000—had started competitive swimming at age five and had accumulated over 7000 h of deliberate practice. However, another elite swimmer did not begin competitive swimming until he was a senior in high school, and had accumulated only about 3000 h of deliberate practice. This late bloomer won Olympic gold after less than *2 years* of serious swimming. Thus, as Macnamara et al. (2015) concluded in their meta-analysis of sports studies, deliberate practice may lose its predictive power at elite skill levels.

In one of the few longitudinal studies of expertise to date, Schneider and colleagues (Schneider, Bös, & Rieder, 1993; Schneider, 1997) tested for effects of a wide range of factors on the development of expertise in elite youth tennis players. (About 10% of the players were ultimately ranked in the top 100 in the world, and a few were rated in the top 10.) The participants completed tests of psychological and physical characteristics, motivation, basic motor abilities, and tennis-specific skills. In addition, biographical interviews were conducted with the players, and their parents and coaches. Measures of competitive tennis success (i.e., ranking) were then obtained for multiple time points. Given the importance and rarity of this type of study, and the high quality of this particular study, we reproduce the structural equation model from the most recent report of the results in Figure 3. As shown, the player's preference for tennis and the coach's rating of future success were strongly predictive of tennis-specific skills, which were strongly predictive of tennis ranking. However, basic motor abilities had an indirect impact on ranking through tennis-specific skills. Schneider thus concluded that "[a]lthough individual differences in basic motor abilities were not large in this highly selected sample, they made a difference when it came to predicting individual tennis performance" (p. 14). Reviewing these and other findings, Schneider (2015) concluded that "whereas Ericsson and colleagues believe that the amount of deliberate practice is a sufficient predictor of

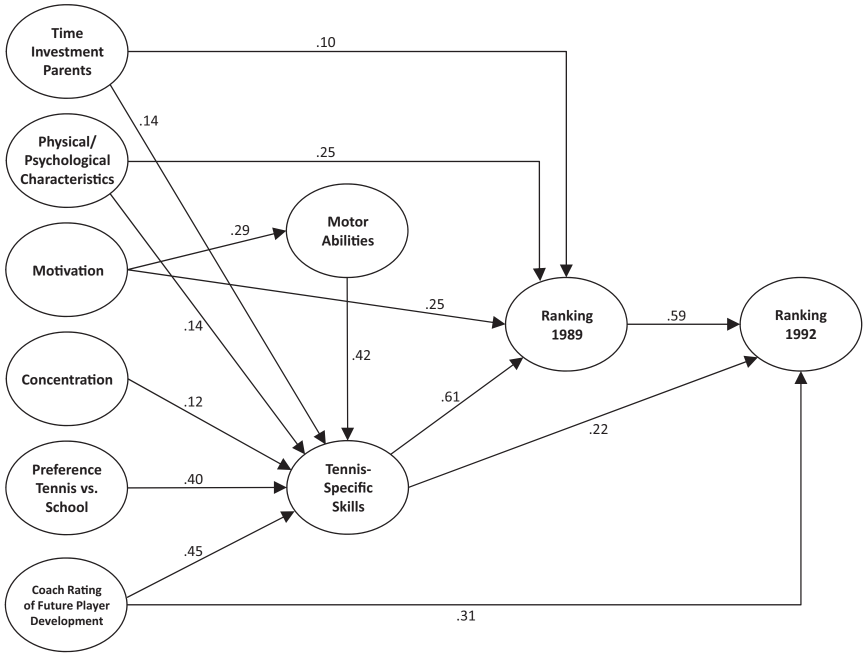

Figure 3 Structural equation model from Schneider (1997) predicting tennis-specific skills and tennis ranking. *Reproduced with permission of Taylor and Francis from Schneider (1997), Figure 5.*

subsequent expert performance, the developmental findings suggest that individual differences cannot be completely ignored when it comes to predicting the development of expertise" (p. 251).

Using a biographical research approach, Lombardo and Deaner (2014) investigated the role of training in athletic success through analyses of biographies and autobiographies of elite sprinters. In one study, Lombardo and Deaner examined the biographies of 15 Olympic gold medalists in the 100-m and 200-m sprints—from Jesse Owens in 1936 to Usain Bolt in 2008 and 2012—and recorded any mention of exceptional (or unexceptional) speed relative to peers. All 15 of the sprinters were recognized as having exceptional speed prior to or from the outset of training. Moreover, the sprinters were found to require between 1 and 7 years to reach world class status, with a mean of 4.6 years ($SD = 2.0$) for the men and 3.1 years ($SD = 2.4$) for the women. In a second study, Lombardo and Deaner used archival records to document the 20 fastest American male sprinters in history. Eight of the 12 sprinters for whom data were available were found to reach world class status in fewer than 10 years ($M = 8.7$, $SD = 3.8$).

These findings are inconsistent with the claim that "winning performances at international competitions within competitive domains of expertise requires more than a decade of preparation" (Boot & Ericsson, 2013, p. 147). At least in sprinting, the 10-year rule does not hold true.

An intriguing case study of deliberate practice and sports expertise is in progress. In April, 2010, having read about Ericsson and colleagues' research, 30-year old Dan McLaughlin quit his job as a commercial photographer, and with virtually no prior experience playing golf, set out to reach the Professional Golfer's Association (PGA) Tour—the highest level of competitive golf in the world—through 10,000 h of deliberate practice. With input from Ericsson and colleagues, McLaughlin worked with golf teaching professionals to design a training regimen based on the concept of deliberate practice (McLaughlin, 2014). McLaughlin regularly records his progress in an online log—the "10,000 hour countdown" (see http://web.archive.org/web/20150803113448/http://thedanplan.com/countdown/), including the number of hours of deliberate practice remaining, the score he shot if he played a round of golf, and qualitative information about his performance. At the 5-year mark, McLaughlin's lowest score for 18 holes was 70, and his lowest handicap (a standardized index of skill level) was 2.6, putting him above the 95th percentile for amateur golfers in the United States (see http://thedanplan.com/).[7]

While McLaughlin's progress is impressive, there are notable examples of people taking up golf relatively late in life (even as adults) and acquiring a much higher level of skill over a 5-year period. In her autobiography, Babe Didrikson Zaharias recalls that she played her first round of golf at age 21 (Zaharias, 1955). Three years later, Zaharias won the Texas Women's Amateur and went on to become one of the greatest golfers in history (Van Natta, 2011). Greg Norman, who was the top-ranked golfer in the world for 331 weeks (see http://www.owgr.com/ranking), recalls in his autobiography that he received his first set of golf clubs at age 15, and soon thereafter recorded his first official score—a 108 (Norman & Phillips, 2006). Just over 3 years later, Norman competed in the Australian Open, and finished with the second lowest score for an amateur and 35th overall. Three years after that, he won his first professional tournament, beating two of the best players in the world at the time. As another example, Larry Nelson took up golf at

[7] For interviews with Dan McLaughlin, K. Anders Ericsson, and others involved in The Dan Plan, see a segment of Golf Channel's *Golf in America* at https://www.youtube.com/watch?v=v4GT0vGS-lA.

age 21. Three-and-a-half years later, he qualified for the PGA Tour, and he has since won 41 professional tournaments, including three major championships (Riach, 2003; Yocom, 2008). Deliberate practice does not appear to be the only factor involved in reaching an elite level of performance in golf, and it may not be the most important factor.

There have also been a few studies of the relationship between deliberate practice and professional expertise. In one of the best to date, Chow and colleagues (Chow, Miller, Seidel, Kane, Andrews, & Thornton, 2015) investigated the impact of deliberate practice on expertise in psychotherapy. The participants were professional psychotherapists, who over a 4-year period asked their more than 1600 clients to complete a questionnaire to assess the effectiveness of their treatment in terms of symptoms, functioning, and risk. The psychotherapists themselves completed a questionnaire in which they estimated the amount of time they spent engaging in activities outside of work to improve therapeutic skills (i.e., deliberate practice). Consistent with previous work (Ericsson et al., 1993), Chow et al. found a statistically significant relationship between average number of hours per week spent alone in deliberate practice and client outcomes. High levels of deliberate practice were associated with lower levels of client distress at the end of therapy. However, even among the therapists with the best client outcomes (the top quartile), there was a large amount of variability in deliberate practice (see Chow et al., Figure 1). Some of the top therapists reported engaging in much more deliberate practice than others.

To sum up, there is now a sizeable body of evidence to indicate that a large amount of variance in expertise is explained by factors other than deliberate practice. To put it another way, in terms of its contribution to individual differences, deliberate practice appears to be an important piece of the expertise puzzle, but only one piece, and not even necessarily the largest piece. What, then, are the other pieces of the puzzle?

4. WHAT ELSE MATTERS?

4.1 Opportunity Factors

Obviously, people are not born with the specialized skills and knowledge that are necessary for success in complex domains such as music and chess. Thus, it stands to reason that people who have a greater opportunity to train in these domains will have an advantage over those who have less of an opportunity to train. As a stark illustration, there are currently over 300 players in Major

League Baseball (MLB) from the Dominican Republic (http://mlb.mlb.com/dr/active_players.jsp?pagina=5)—more than any other country in the world except the United States—and none from Haiti, which borders the Dominican Republic on the island of Hispaniola. The major reason for this difference is almost certainly opportunity: baseball is a national priority in the Dominican Republic (Klein, 1993), but not in much poorer Haiti.

Nationality is an example of an "opportunity" factor that would be expected to impact expertise indirectly, through deliberate practice and other forms of training. Parental influence is another example. Bloom and colleagues interviewed highly accomplished musicians, artists, athletes, and academics to better understand the origins of their success (Bloom, 1985). The overall conclusion of the study was that "no matter what the initial characteristics (or gifts) of the individuals, unless there is a long and intensive process of encouragement, nurturance, education, and training, the individuals will not attain extreme levels of capability in these particular fields" (Bloom, 1985, p. 3).

Birth date is another example of an opportunity factor. For some sports, such as hockey, there is some evidence that individuals born early in the year have a greater chance of reaching the professional ranks than individuals born later in the year (Barnsley, Thompson, & Barnsley, 1985). One proposed explanation of these *relative age effects* is that players born near the eligibility cutoff for participation at a given age level (e.g., in a league) will be older and physically more mature and capable than players with a later birth date, and thus will be singled out as "talented" and given more opportunities to train and acquire expertise.

4.2 Basic Ability Factors

Some people acquire complex skills much more rapidly than other people. Consider that Magnus Carlsen achieved grandmaster status—the highest possible rating in tournament chess—at age 13, less than 5 years after competing in his first chess tournament (Agdestein, 2013). Or consider that Donald Thomas won his first collegiate high jump competition with almost no training in the event (Epstein, 2014), and within two years competed in the Olympics. Cases such as these raise the question of whether people differ in the basic abilities—*talents*—that they can bring to bear on acquiring expertise.

We have focused on the role of *working memory capacity* (WMC) as a form of intellectual talent. WMC is the ability to maintain information in an active and accessible state over a short period of time (Engle, 2002) and is measured with tasks such as *operation span*, in which the participant attempts to solve arithmetic equations while simultaneously remembering words. WMC correlates moderately with performance in a wide range of complex cognitive tasks, including text comprehension, decision making, and reasoning (Hambrick & Engle, 2003). Heritability estimates for WMC are usually around 50% (e.g., Ando, Ono, & Wright, 2001; Kremen et al., 2007; Polderman et al., 2006).

Consistent with classical models of skill acquisition (e.g., Anderson, 1982; Fitts & Posner, 1967), Ericsson and colleagues have argued that WMC and other basic abilities impact performance only initially during training, after which their influence is circumvented through specialized knowledge and skills that develop through deliberate practice. As Ericsson and Charness (1994) stated, "[t]he effects of extended deliberate practice are more far-reaching than is commonly believed. Performers can acquire skills that circumvent basic limits on working memory capacity and sequential processing" (p. 725). And as Ericsson (2014a) reiterated, "[t]he acquisition of expert performance, where acquired mechanisms gradually circumvent the role of any basic general cognitive capacities and thus reduce and even eliminate significant relations between general cognitive ability and domain-specific performance at the expert level of performance" (p. 83).

Though they did not explicitly frame it as such, Robbins et al. (1996) tested this *circumvention-of-limits hypothesis* using an experimental approach. Chess players, ranging in skill from "weak club player" to master, performed a move-choice task while performing secondary tasks designed to suppress various components of the working memory system, or with no secondary task (the control condition). Robbins et al. found that a secondary task designed to tap the *central executive* component of working memory—the domain-general system responsible for higher-level cognitive processes (Baddeley & Hitch, 1974)—was severely disruptive to participants' performance in the move-selection task, regardless of skill level. A secondary task designed to tap the *visuospatial sketchpad* was similarly disruptive. These results suggest that working memory directly influences performance in chess. More recently, Foroughi, Werner, Barragán, and Boehm-Davis (2015) found that interruptions designed to clear the transient working memory were disruptive to reading comprehension in skilled readers.

We have used an individual differences approach to evaluate the circumvention-of-limits hypothesis. As illustrated in the left panel of Figure 4 the prediction that follows from this hypothesis is an interaction between a domain-general factor (e.g., WMC) and a domain-specific factor (e.g., deliberate practice) on domain-specific performance. That is, at high levels

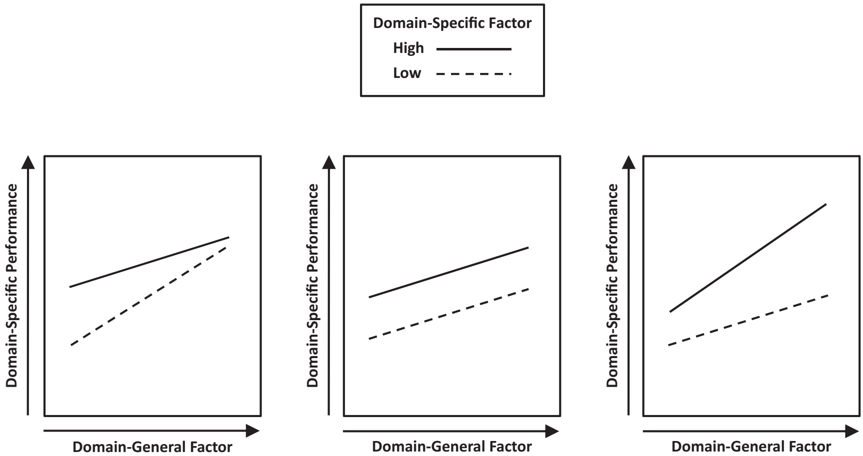

Figure 4 Three hypotheses concerning effects of domain-general and domain-specific factors on domain-specific performance (expertise): *circumvention-of-limits hypothesis* (left panel), *building blocks hypothesis* (middle panel), and *rich-get-richer hypothesis* (right panel).

of the domain-specific factor (e.g., deliberate practice), the domain-general factor (e.g., WMC) is less predictive of performance than at lower levels of the domain-specific factor. There are two alternative hypotheses (see also Hambrick & Engle, 2002). The *building blocks hypothesis* (middle panel) predicts additive effects of the domain-general and domain-specific factors on performance; that is, the effects of domain-general factors on performance are statistically equivalent across levels of the domain-specific factors. The *rich-get-richer hypothesis* (right panel) predicts a domain-general × domain-specific interaction, but in the opposite direction to that predicted by the circumvention-of-limits hypothesis: a stronger effect of the domain-general factor at high levels of the domain-specific factor.

To test these possibilities, Meinz and Hambrick (2010) had 57 pianists, ranging in skill from beginner to professional, complete a questionnaire to assess deliberate practice, along with tasks to measure both WMC and sight-reading ability. (Sight-reading involves playing music with little or no preparation.) Deliberate practice accounted for nearly half (45%) of the variance in sight-reading performance, but WMC accounted for an additional 7% of the variance. More important, there was no deliberate practice × WMC interaction. Instead, consistent with the building-blocks hypothesis, the effect of WMC on performance was as large at low levels of deliberate practice as at higher levels of deliberate practice. For all but the most difficult piece of music they used in their study, Kopiez and Lee (2006) also reported significant positive correlations between a measure of working memory and sight-reading performance (see Hambrick & Meinz, 2012; for a review of music studies). Furthermore, in a study of Texas Hold'Em poker, Meinz et al. (2012) found that WMC positively predicted performance in poker skill tasks (e.g., hand evaluation), even at high levels of poker knowledge. Similarly, Toma et al. (2014) found that both SCRABBLE and crossword experts outperformed control subjects on two tests of WMC.

Research on prodigies lends further support to the conclusion that WMC plays an important role in acquiring expertise. Ruthsatz and Detterman (2003) documented the case of a 6-year old piano prodigy ("Derek") who had played in numerous concerts, appeared on national television, and released two CDs of his music. Derek scored at or above the 95th percentile on tests of musical aptitude. He also scored well above the average on the verbal reasoning (130), abstract reasoning (114), and quantitative reasoning (120) subsets of the Stanford-Binet Intelligence Scale, and above the 99th percentile on the short term memory subtest (158). More recently, Ruthsatz and Urbach (2012) administered a standardized IQ test (the Stanford-Binet) to eight child prodigies, six of whom were musical prodigies. Despite full-scale IQs that ranged from 108 to 147—just above average to exceptional—all of the prodigies were at or above the 99th percentile for working memory (indeed, six scored at the 99.9th percentile). Adding nine prodigies to the sample (for a total *N* of 17), Ruthsatz and colleagues found an average score of 140 ($SD = 11.8$) for working memory—2.5 standard deviations above the mean (Ruthsatz, Ruthsatz-Stephens, & Ruthsatz, 2014).

Taken together, this evidence suggests that there are conditions under which WMC limits the ultimate level of performance a person can achieve in a domain. This is not to say that there are *no* conditions under which WMC and other basic abilities can be circumvented. Hambrick et al. (2012) found that visuospatial ability predicted success in a geological bedrock mapping task in which the goal was to infer the geological structure of an area based on observable features (rock outcrops, topography, etc.), but only in participants with low levels of geological knowledge. Similarly, in a study of pilots, Sohn and Doane (2004) found that WMC predicted success in an aviation situational awareness task, but only in pilots who scored low on an aviation-specific test measuring skilled access to long-term memory (i.e., long-term working memory; Ericsson & Kintsch, 1995). For pilots who scored high on this test, there was no relationship between WMC and performance in the situation-awareness task.

As we have noted elsewhere (Hambrick & Meinz, 2011b), this mixed evidence for the circumvention-of-limits hypothesis suggests that there may be task and situational factors that moderate the interplay between domain-general and domain-specific factors. For example, in contrast to domains in which the stimulus input is static (e.g., geological bedrock mapping), tasks in which the input changes continuously and rapidly and is unpredictable (e.g., sight-reading) may make it more difficult to rely on long-term memory knowledge structures to circumvent WMC and other basic abilities. Admittedly, this is a posthoc speculation, and as we discuss later, a goal for future research is to develop a framework for making testable predictions about how task/situational factors impact expertise.

Numerous other studies have investigated the relationship between expertise and traditional measures of IQ and specific cognitive abilities (verbal ability, visuospatial ability, etc.). This research has tended to suffer from serious methodological limitations (see Ackerman, 2014), including use of (1) extremely small sample sizes, leading to low statistical power and precision; (2) samples with restricted ranges of cognitive ability and/or expertise, limiting the degree to which the variables can correlate; (3) single tests of cognitive ability, leaving open the question of whether results are test-specific; (4) tests with unknown reliability and validity; and (5) research designs that confound skill level (e.g., novice vs expert) with other factors that may account for group differences in cognitive ability (e.g., educational status). Further complicating matters, participants are sometimes children, and other times adults.

Not surprisingly, then, the results of these studies concerning ability–expertise relations are inconsistent (see Ericsson, 2014a; for a review). Whether in terms of correlations or differences in group means, relationships between cognitive ability and expertise are sometimes found to be statistically significant and sizeable, and other times not. A comprehensive review of this literature is beyond the scope of this chapter. A few examples will suffice to illustrate the inconsistency. Using small samples of tournament chess players as participants, Unterrainer et al. (Unterrainer, Kaller, Halsband, & Rahm, 2006; Unterrainer, Kaller, Leonhart, Rahm, 2011) reported nonsignificant correlations between IQ and chess rating: $rs = -0.08$ ($N = 25$) and -0.07 ($N = 30$), respectively. However, using a considerably larger sample ($N = 90$), Grabner, Stern, and Neubauer (2007) found a correlation of 0.35 ($p < 0.001$) between IQ and chess rating.

One study on the relationship between cognitive ability and expertise stands out as methodologically superior: Masunaga and Horn's (2001) study of GO expertise. What do the results of this study suggest? In this study, participants ($N = 263$) representing wide ranges of age, cognitive ability, and expertise in the board game GO completed tests of both domain-general and domain-specific cognitive abilities. The domain-general battery included standard tests of fluid reasoning (Gf), short-term memory (Gsm), and perceptual speed (Gs), whereas the domain-specific battery included "GO-embedded" tests. The GO-embedded tests were designed to measure Gf, Gsm, and Gs, but with GO-specific content. Particularly relevant to the present discussion, the GO reasoning test was explicitly modeled after tasks used to objectively measure skill in chess (e.g., de Groot, 1946/1978). The participants were given GO game positions and asked to choose the next best move. The best answers in this task were determined by GO professionals (see Masunaga's, neé Takagi, 1997, dissertation for additional information on the development of the task).

Figure 5 presents a reanalysis of Masunaga and Horn's (2001) published results using structural equation modeling (SEM; from Hambrick & Macnamara, 2016). (All that is required for SEM is a correlation matrix

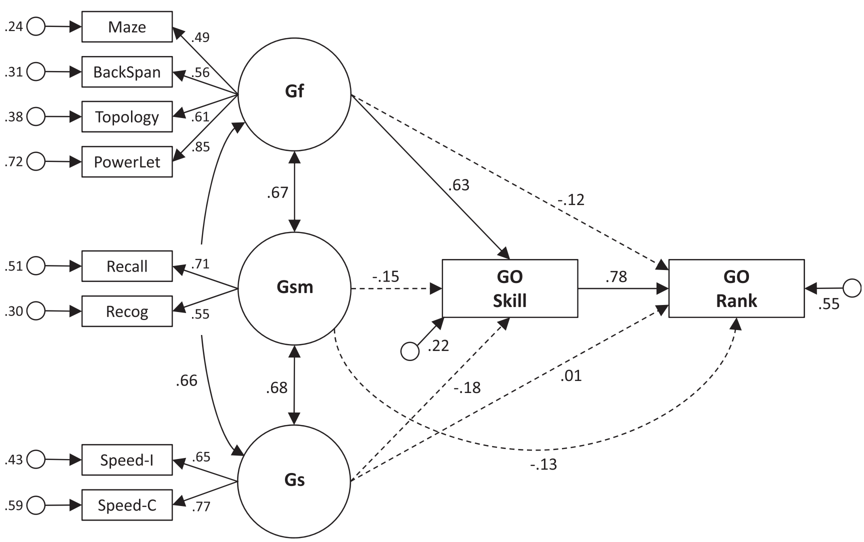

Figure 5 Reanalysis of published results of Masunaga and Horn (2001), with domain-general cognitive abilities (Gf, Gsm, and Gs) predicting GO skill and GO rank. Values adjacent to single-headed arrows are standardized path coefficients; values adjacent to double-headed arrows are correlations. Solid paths are statistically significant ($p < 0.01$). Correlations for reanalysis obtained from Masunaga and Horn's Tables 6, 9, and 10. Model fit is excellent: $\chi^2(27) = 28.45$, $p = 0.39$, CFI = 1.0, NFI = 0.96, RMSEA = 0.01. $R^2 = 0.22$ for GO skill and 0.55 for GO rank.

among the variables of interest, which Masunaga and Horn provided.) One of the major advantages of SEM over other statistical approaches is that it permits analysis of data at the level of *latent variables* (see Kline, 2011; for an excellent introduction to SEM). A latent variable captures variance common to multiple observed variables, and thus statistically cancels out task-specific factors and random measurement error. The purpose of the SEM reanalysis shown in Figure 5 was to test for effects of latent variables representing the domain-general abilities on GO skill, as measured by the GO reasoning task, and on GO ranking. As shown, domain-general Gf was positively predictive of GO skill (0.63, $p < 0.001$): high levels of Gf were associated with high levels of GO skill. In turn, GO skill was positively predictive of GO rank (0.78, $p < 0.001$). (These relationships were very similar after statistically controlling for age in the model: 0.54 and 0.76, respectively, $ps < 0.001$.) This evidence suggests that domain-general Gf contributes to individual differences in the type of task that the expert performance approach requires for use in expertise research (Ericsson & Smith, 1991; Boot & Ericsson, 2013).

Grabner, Stern, and Neubauer's (2007) study of chess expertise is also worthy of further discussion, given that the study used a relatively large sample with wide ranges of both expertise and cognitive ability, and multiple tests of cognitive ability with established reliability and validity. The participants (chess rating = approximately 1300 to 2400, or novice to master) completed a standardized test of intelligence, with numerical, verbal, and figural subscales. The sample was approximately one standard deviation above the mean for the general population in general intelligence (i.e., $M = 114$, $SD = 14$). Moreover, chess rating correlated moderately and positively with general intelligence ($r = 0.35$), and both numerical intelligence ($r = 0.46$) and verbal intelligence ($r = 0.38$). (The correlation with figural intelligence was near zero.) From their data, Grabner et al. estimated the minimum verbal and numerical IQ necessary to achieve an "expert" or "advanced" status (Elo rating > 2200) to be between 110 and 115 (or 0.67 and 1 *SDs* above the mean of the general population). For full-scale IQ, the lowest IQ for a player with an Elo rating above 2200 was about 103.[8]

The results of the landmark Study of Mathematically Precocious Youth are also relevant (see Robertson, Smeets, Lubinski, & Benbow, 2010).

[8] We thank Roland Grabner for e-mailing us a scatterplot from this study showing the correlation between full-scale IQ and chess rating (personal communication, May 6, 2015).

As part of a youth talent search, a large sample of children took the SAT by age 13, and those scoring in the top 1% ($N > 2000$) were identified and tracked over the next two decades. Even within this group, SAT score predicted individual differences in objective measures of educational and professional accomplishment. For example, compared to participants in the 99.1 percentile for overall SAT score, participants who had scored in the 99.9 percentile were 3.6 times more likely to have earned a doctorate, 5 times more likely to have published an article in an STEM journal, and 3 times more likely to have registered a patent (Lubinski, 2009). More recently, Lubinski, Benbow, and Kell (2014) found that accomplishments of intellectually talented individuals (top 1% for mathematical reasoning) far exceeded base-rate expectations. For example, 2.3% of the sample were CEOs at major companies, and 4.1% had earned tenure at a major research university.

Cognitive ability does not *always* predict individual differences in expertise. With a sample size of over 700, Lyons, Hoffman, and Michel (2009) analyzed data from the National Football League's (NFL) Combine, a weeklong event in which players who aspire to play in the NFL demonstrate their skills and perform various tests of physical and mental ability. Lyons et al. found that scores on a standardized test of cognitive ability (the Wonderlic Personnel Test) generally correlated near zero with success in the NFL across all positions considered. Berri and Simmons (2011) performed a more detailed analysis of the performance of quarterbacks, and once again found no evidence that Wonderlic scores predicted future NFL performance. Football may thus be a domain in which cognitive ability does not play any appreciable role in success. Alternatively, it could be that cognitive abilities not captured by the Wonderlic, such as WMC, perceptual speed, and psychomotor speed predict performance in football, or that team-level factors override the impact of individual-level factors.

To summarize, there is consistent and compelling empirical evidence that cognitive ability predicts individual differences in expertise in some, if not all, domains. Ericsson has reached a different conclusion in his own reviews. Ericsson, Prietula, and Cokely (2007) concluded that "there is no correlation between IQ and expert performance in fields such as chess, music, sports, and medicine" (p. 116) and that the "only innate differences that turn out to be significant—and they matter primarily in sports—are height and body size" (p. 116). And in a more recent review, Ericsson (2014a) concluded:

> *Let it be clear that I am not claiming that correlation between domain-specific performance and general cognitive ability is exactly zero!! My current conclusion is that*

> *these studies have* not yet *established the fact that the attainable level of domain-specific performance is predictable from scores from tests of general cognitive ability.* (p. 87)

However, as we and others have noted (Ackerman, 2014; Hambrick, Altmann, et al., 2014), Ericsson appears to overlook evidence that contradicts this conclusion. For example, in his most recent review, Ericsson (2014a) mentioned Meinz and Hambrick's (2010) study of piano sight-reading, but he did not mention the central result of this study—that there was no interaction between WMC and deliberate practice, indicating that WMC was as predictive of sight-reading performance at low levels of deliberate practice as at high levels. As another example, although Ericsson correctly noted that the domain-general cognitive ability measures correlated near zero with GO rating in the Masunaga and Horn (2001) study, he does not mention the fact that nearly all of the other correlations between these cognitive ability measures and the GO-embedded measures were statistically significant (i.e., $p < 0.01$ for 50 out of 56 of the *r*s).

Moreover, Ericsson (2014a) makes material errors in his review (see Hambrick, Altmann, et al., 2014). This is understandable, particularly given the scope of his review. All the same, these errors are serious enough that they could lead to significant confusion if the scientific record is not corrected. One material error directly relevant to this discussion is Ericsson's claim that Grabner et al. (2007) "report that one chess master with a rating close to 2400 had an IQ of around 80" (Ericsson, 2014a, p. 87). If true, this would be somewhat surprising. A person with an IQ of around 80 (the 9th percentile for the general population) falls in the range for what is sometimes referred to as *borderline intellectual functioning* (see Peltopuro, Ahonen, Kaartinen, Seppälä, & Närhi, 2014). However, Grabner et al. reported no such result. There was one player with a rating close to 2400 and a *figural* IQ of 70, indicating that this individual had low scores on the figural reasoning subtests. However, this same player had a numerical IQ of 117 (the 87th percentile), a verbal IQ of 113 (the 81st percentile), and a full-scale IQ of 103 (the 58th percentile; Roland Grabner, personal communication, May 6, 2015). There is no report in Grabner et al.'s article of a chess master with a rating close to 2400 and an IQ of around 80.

Ericsson (2014a) also makes points concerning the relationship between cognitive ability and expertise that do not stand to reason. For example, he notes that Garry Kasparov's IQ was estimated at 120 based on Raven's Progressive Matrices (Der Spiegel, 1987), "which is very close to the average of all chess players…thus not very predictive of world-class chess performance"

(Ericsson, 2014a, p. 87). However, one case does not a correlation make: if Kasparov was an outlier, and other world champion chess players (Boris Spassky, Bobby Fischer, Magnus Carlsen, etc.) had extremely high IQs, then IQ could still be *highly* predictive of world-class chess performance.

Thus, although we credit Ericsson (2014a) for his review—it will be essential reading for anyone interested in expertise for years to come—we disagree with his claim that there is currently no evidence to suggest that cognitive ability significantly predicts expertise. To be sure, correlations between cognitive ability and expertise are often not as large as those between deliberate practice and expertise, but neither are they trivially small, from either a statistical or a practical perspective. This conclusion is broadly consistent with evidence that cognitive ability is the single best predictor of job performance and maintains its predictive validity even in highly experienced employees (see Schmidt & Hunter, 2004). It also falls in line with Ackerman's (2014) observation that "there is ample evidence from over 100 years of research supporting the conclusion that abilities are significantly related to individual differences in the attainment of expert performance" (p. 104).

4.3 Personality Factors

A central theme of the biographies of many elite performers is intense commitment to their domains—a singular devotion seeming to border on the pathological. As a student, Marie Curie frequently forgot to eat, and even after winning her first Nobel Prize, she would work in her lab past midnight (Goldsmith, 2005). The golfer Ben Hogan is said to have hit practice balls until his hands bled and then soaked his blistered hands in pickle brine to toughen them so he could practice more (Dodson, 2005). Winner (2000) described such focus as a "rage to master," and noted that children who possess this quality "have a powerful interest in the domain in which they have high ability, and they can focus so intently on work in this domain that they lose sense of the outside world" (p. 162; see also Winner, 1996a, 1996b).

Ericsson et al. (1993) hypothesized that a number of personality factors predispose people to intense commitment to their domain:

> *within our framework we would expect that several 'personality' factors, such as individual differences in activity levels and emotionality may differentially predispose individuals toward deliberate practice as well as allow these individuals to sustain very high levels of it for extended periods.*
> *(p. 393)*

This view leads to the prediction that deliberate practice should mediate the effect of personality factors on domain-specific performance. There is support for this prediction. In a study of Spelling Bee contestants, Duckworth, Kirby, Tsukayama, Berstein, and Ericsson (2012) found that "grit"—a personality factor reflecting persistence in accomplishing long-term goals (Duckworth & Gross, 2014)—positively predicted deliberate practice, which in turn positively predicted spelling performance. Along the same lines, in a study of classical musicians, Bonneville-Roussy, Lavigne, and Vallerand (2011) found that "passion" positively predicted "mastery goals," which positively predicted deliberate practice, which positively predicted music performance (see Vallerand, 2015; see also Hallam, 1998). Similarly, in a study of chess players, de Bruin, Rikers, and Schmidt (2007) found that a measure of motivation to engage in deliberate practice positively predicted accumulated amount of deliberate practice, which in turn positively predicted chess rating.

This evidence supports the idea that people differ in their propensity to engage in deliberate practice, which translates into individual differences in expertise. However, personality factors may also impact performance directly. For example, Grabner et al. (2007) found that chess rating correlated positively with a measure of the ability to regulate the expression of emotions, even after controlling for a number of other factors (intelligence, number of tournament games, motivation, etc.). High levels of emotional control were associated with superior chess skill. Susceptibility to performance anxiety and to "choking" under pressure are other personality-type factors that could impact performance directly, independent of deliberate practice.

4.4 Other Domain-Relevant Experience Factors

Experts spend a considerable amount of time training, but obviously they engage in other forms of domain-relevant experience as well. Ericsson et al. (1993) distinguished deliberate practice from two other types of domain-relevant activities, which they termed *work* and *play*, as follows:

> *Work includes public performance, competitions, services rendered for pay, and other activities directly motivated by external rewards. Play includes activities that have no explicit goal and that are inherently enjoyable. Deliberate practice includes activities that have been specially designed to improve the current level of performance. (p. 368)*

The deliberate practice view claims that these other forms of domain-relevant experience are weaker predictors of domain-specific performance than deliberate practice. As Boot and Ericsson (2013) explained, "Ericsson and colleagues...make a critical distinction between domain-related activities of work, play, and deliberate practice, and claim that the amount of accumulated time engaged in deliberate practice activities is the primary predictor of exceptional performance" (p. 146).

This claim leads to the prediction that measures of deliberate practice should correlate more strongly with expertise than measures of engagement in either work or play. There is some evidence to support this prediction. For example, in their two studies, Charness et al. (2005) found that log hours of tournament play (work) did not significantly predict chess rating after controlling for log hours of serious study (deliberate practice). However, this prediction is not always supported. Howard (2012) found total number of games correlated almost twice as strongly with chess rating as total study hours did, and in a study of insurance salespeople, Sonnentag and Kleine (2000) found that the number of cases handled—a measure that fits the description of work—correlated more strongly with a measure of sales performance ($r = 0.37$) than measures of both current and accumulated deliberate practice did (rs $= 0.21$ and 0.13). As another example, Moxley, Ericsson, Scheiner, and Tuffiash (2015) found that log number of years of participating in crossword puzzle tournaments correlated significantly with performance in the American Crossword Puzzle Tournament ($r = 0.32$). *A priori,* participating in a tournament would seem to be a clear instance of what Ericsson et al. (1993) described as work.

Other studies have found that play positively predicts performance. For example, Ford and Williams (2012) found that youth soccer players who had received professional scholarships at age 16 had engaged in significantly more soccer play-like activities per year than the soccer players at the same clubs who had been asked to leave at age 16 for not making significant progress. Haugaasen, Toering, and Jordet (2014) found similar results with youth soccer players: those who had received professional contracts had engaged in more play activities during early development (ages 6–8) than their soccer club counterparts who had not received professional contracts. On a related note, Côté and colleagues have found that *deliberate play,* which they define as activities that are "intrinsically motivating, provide immediate gratification, and are specifically designed to maximize enjoyment" (Côté, Baker, & Abernethy, 2007, pp. 185–186), can be as predictive of expertise as deliberate practice (see, e.g., a study of ice hockey by Soberlak & Côté, 2003).

To sum up, there is evidence that forms of domain-relevant experience other than deliberate practice, including what Ericsson et al. (1993) termed work and play, significantly predict expertise, and are perhaps even more predictive than deliberate practice in some domains.

4.5 Developmental Factors

For the obvious reason that expertise in virtually all domains is acquired gradually, a complete account of the origins of expertise must take into account developmental factors. One developmental factor is *starting age*. Reviewing evidence from a small number of studies, Ericsson et al. (1993) concluded that "we find that the higher the level of attained elite performance, the earlier the age of first exposure as well as the age of starting deliberate practice" (p. 389) and "[a]cross many domains of expertise, a remarkably consistent pattern emerges: The best individuals start practice at earlier ages and maintain a higher level of daily practice" (p.392). Ericsson et al. further argued that the benefit of starting early (vs later) is a longer period of time to accumulate deliberate practice: "[t]he individuals who start early and practice at the higher levels will have a higher level of performance throughout development…than those who practice equally hard but start later" (p. 392).

This argument leads to the prediction that the effect of starting age on performance should be mediated through deliberate practice. Consistent with this hypothesis, in an initial report of data from their study of chess players (reported in Charness, Krampe, & Mayr, 1996), Charness and colleagues found that the relationship between starting age and chess rating was nonsignificant after statistically controlling for accumulated amount of deliberate practice. However, both Gobet and Campitelli (2007) and Howard (2012) found that the correlation between starting age and chess rating was statistically significant even after statistically controlling for accumulated amount of deliberate practice. This evidence is consistent with the possibility that there is a critical period for acquiring some complex skills, just as there may be for language.

A complete account of expertise must also take into account effects of aging. Though it is clear that various aspects of physical, sensory, perceptual, motor, and cognitive functioning decline in adulthood (Santrock, 2012), findings from cross-sectional research on aging and expertise are inconsistent. For example, although Masunaga and Horn (2001) found a near zero correlation between age and GO ranking among amateur players ($r = 0.04$), Moxley and Charness (2013) found an average correlation of −0.28 between age and performance in best move tasks in chess. One possible explanation for this inconsistency is selective attrition; that is, weak performers may quit. A more consistent pattern of results emerges from longitudinal studies: performance increases up to a peak age, after which it decreases. In intellectual domains, the peak age tends to be in the mid-30s to mid-40s. For example, in a longitudinal analysis of over 5000 chess players, Roring and Charness (2007) found a peak age of 43.8 years for chess rating, and Simonton (1991) documented peak ages (i.e., age of best contribution) in the mid-30s to early-40s for academic domains. In physical domains, the peak age is much earlier. For example, Schulz and Curnow (1988) found that the average age of Olympic gold medal winners is in the early 20s for short-distance running events (e.g., 22.9 years for the 100 m) and the late 20s for long-distance events (e.g., 27.9 for the marathon). It has been suggested that age-related decline in skill is not inevitable and instead reflects reduction in deliberate practice (Krampe & Ericsson, 1996), but at present, there is very little evidence to support this hypothesis (see Hambrick & Macnamara, 2016).

4.6 Genetic Factors

The finding that (1) deliberate practice leaves a large amount of individual differences in expertise unexplained and (2) basic abilities known to be influenced by genetic factors correlate with expertise in these same domains, suggests that individual differences in genetic factors also contribute to individual differences in expertise. However, this evidence is merely suggestive of a genetic contribution, for the obvious reason that these same basic abilities are also known to be influenced by *environmental* factors. More direct evidence for an impact of genetic factors on expertise comes from behavioral genetics research.

Although it is difficult to quantify the degree to which two people's environments are similar, it is relatively easy to quantify the degree to which they share genetic factors. This is because inheritance of most genetic material follows very simple rules, which were first postulated by Gregor Mendel in the mid-1800s based on his experiments with pea plants (Mendel, 1866). Biometrical theory can be used to calculate the average amount of genetic sharing between two relatives at the genome-wide level. Like siblings, a child shares 50% of their autosomal (i.e., non-sex chromosome) DNA with each of their parents. By contrast, grandparents share on average 25% of their genetic material with their grandchildren (like half-siblings and members of avuncular relationships). Making use of this information about differences in average genetic sharing between relatives, analysis of data from related individuals (the family design) enables statistical estimation of the relative magnitude of genetic and environmental influences on trait variation (Blokland, Mosing, Verweij, & Medland, 2013).

The twin study is the most commonly used family design, and compares within-pair similarity of identical (monozygotic; MZ) and non-identical (dizygotic; DZ) twins. MZ twins are genetically identical, whereas DZ twins share on average only 50% of their genetic loci. However, both types of pairs have shared prenatal environments (as they were conceived at the same time and shared the womb) and also share much of their rearing environment, as they are born at approximately the same time and grow up together in the same family environment. Such environmental influences common to the two members of a twin pair are generally referred to as shared environmental influences and are assumed to make the twins more similar to each other. Finally, there are also environmental influences that are unique to each one of the twins and will make the members of a twin pair more different from each other (e.g. a trauma, different friends or teachers). Such influences are referred to as non-shared environmental influences. In twin modeling, the non-shared environmental estimates will also include measurement error. Via SEM, genetic versus environmental influences on the variance in a trait can be disentangled and quantified. Heritability refers to the proportion of the phenotypic variance in a trait that is attributable to the effects of genetic variation (Neale & Cardon, 1992).

Twin research has now convincingly established that observed (or *phenotypic*) differences in complex human traits are influenced by both genetic and environmental factors, including their interaction and correlation (Polderman et al., 2015). For example, heritability estimates typically range from 50% to 70% for general intelligence, and from 30% to 50% for specific cognitive abilities and personality traits (Plomin, DeFries, McClearn, & McGuffin, 2008). Given that these same factors appear to play an important role in expertise, it is reasonable to also expect genetic influence on variation in expertise (Bouchard & McGue, 2003; Harris, Vernon, Johnson, & Jang, 2006; Plomin & Spinath, 2004), and there is evidence that this is the case. Coon and Carey (1989) used a sample of over 800 same-sex twin pairs to investigate the heritability of music accomplishment. The twins in this sample were identified through a survey given to roughly 600,000 high school juniors who took the National Merit Scholarship test in 1962 (see Loehlin & Nichols, 1976). The twins completed a survey to determine whether they were identical or fraternal, and then completed a 1082-item psychosocial survey that included several questions about both music accomplishment and music practice. For a measure of musical achievement, the heritability estimate was 38% for males and 20% for females. Vinkhuyzen, van der Sluis, Posthuma, and Boomsma (2009) analyzed data from a study in which 1685 twin pairs rated their competence in chess, music, and several other domains and found even stronger evidence for a role for genetic factors. For endorsement of "exceptional talent," heritability ranged from 50% to 92%.

More recently, in a large sample of adolescent twins, Plomin and colleagues found that genetic factors accounted for over half of the variation between expert and less skilled readers, where experts were defined as individuals who scored above the 95th percentile on a standardized test of reading ability (Plomin, Shakeshaft, McMillan, & Trzaskowski, 2014). Similarly, moderate heritability estimates for objective measures of music ability have been reported in a number of studies. Drayna, Manichaikul, de Lange, Snieder, and Spector (2001) reported heritability estimates of 80% for performance on the Distorted Tunes Test, which requires the participant to identify incorrect pitches from familiar melodic stimuli. Ullén, Mosing, Holm, Eriksson, and Madison (2014) had a sample of over 10,000 twins complete a test of musical aptitude (the Swedish Musical Discrimination Test) and found heritability estimates of 50% for rhythm discrimination, 59% for melody discrimination, and between 12% and 30% for pitch discrimination. Evidence further suggests that the predisposition to practice seems to be partly heritable: in two studies, heritability estimates were between 38% and 70% for music practice (Hambrick & Tucker-Drob, 2014; Mosing, Madison, Pedersen, Kuja-Halkola, & Ullén, 2014).

Behavior genetic studies have also shed light on the nature of phenotypic associations between practice and performance. Mosing et al. (2014) found that both music practice and music aptitude were substantially heritable and that *genetic pleiotropy*—which occurs when one gene or set of genes influences multiple traits—explained much of the association between these two factors. Furthermore, the results of monozygotic intrapair difference modeling, using a co-twin control design, did not support a causal effect of music practice, for either music discrimination (Mosing et al., 2014) or for accuracy of motor timing (Ullén, Mosing, & Madison, 2015). That is, identical twins differing massively in accumulated amount of practice (as much as 20,000 h) performed similarly on tests of these abilities. Similarly, using the National Merit Twin Sample, Hambrick and Tucker-Drob (2014) reported common genetic effects on music practice and music accomplishment. However, the genetic effect on music practice explained only about a quarter of the genetic effect on music accomplishment. This finding is inconsistent with the idea that genetic effects on expertise are entirely mediated through factors that predispose people to engaging in practice. Hambrick and Tucker-Drob also found that the importance of genetic factors for musical accomplishment *increased,* rather than decreased, with practice. This evidence runs counter to Ericsson's (2007) claim that training activities "selectively activate dormant genes that all healthy children's DNA contain" (p. 4), and instead suggests that training may activate dormant genes, variants of which *differ* across individuals. This evidence is in line with an earlier twin study on training of the rotary pursuit task, which found that genetic influences on performance as well as learning rate increased after 3 days of training (Fox, Hershberger, & Bouchard, 1996).

Taken together, findings of these twin studies indicate that there are both direct and indirect effects of genetic factors on expertise. More specific information about the role of genetic factors in expertise comes from molecular genetics research, which seeks to identify associations between specific genes and performance. In a series of studies, North and colleagues documented correlations between genotype for the ACTN3 gene, which codes the alpha-actinin-3 protein in fast-twitch muscles, and performance in various sprint events. For example, in one study (Yang et al., 2003), compared to 18% of control subjects, only 6% of 107 elite athletes from various short-distance events had a variant of ACTN3 that made them alpha-actinin-3 deficient. Even more striking, *none* of the most elite athletes in the sample—the 32 Olympians—were alpha-actinin-3 deficient.

There is also an emerging molecular genetic literature on music (see Tan, McPherson, Peretz, Berkovic, & Wilson, 2014; for a review). In a very recent study, Di Rosa and colleagues (Di Rosa, Cieri, Antonucci, Stuppia, & Gatta, 2015) used Ingenuity Pathway Analysis (IPA) to identify possible interactions between genes potentially related to musical ability and those deleted in individuals with Williams Syndrome—a genetic disorder that is associated with serious deficits in some cognitive domains but surprisingly good musical skills. Di Rosa et al. reported a potential interaction between a gene related to Williams Syndrome (STX1A) and one related to music skills (SLC6A4). Both of these genes are involved in serotonin transporter expression, suggesting that serotonin may be involved in the development of musical abilities.

5. TOWARD A COMPREHENSIVE MODEL OF EXPERTISE

Nearly 20 years ago, Simonton (1999) urged psychologists to think broadly about the potential causes of expertise:

> *it is extremely likely that environmental factors, including deliberate practice, account for far more variance in performance than does innate capacity in every salient talent domain. Even so, psychology must endeavor to identify all of the significant causal factors behind exceptional performance rather than merely rest content with whatever factor happens to account for the most variance.*
> *(p. 454)*

More recently, Kaufman (2014) noted, "Other traits beyond deliberate practice are critical for the development of expert performance." Here, we have discussed a wide range of factors that may contribute to individual differences in expertise. How can all these factors be investigated in an integrative fashion?

5.1 Existing Theoretical Models to Guide Research on Expertise

There exist a number of theoretical frameworks that can guide integrative research on expertise (see Subotnik, Olszewski-Kubilius, & Worrell, 2011; for a comprehensive review). One of the most prominent is Gagné's *Differentiating Model of Giftedness and Talent* or DMGT (Gagné, 2009, 2013, 2014). The DMGT describes how outstanding levels (top 10%) of genetically influenced abilities ("gifts") are transformed into outstanding levels of knowledge and skill ("talents") in occupational fields. DMGT posits that talent development is moderated by "catalysts"—intrapersonal factors such as physical and mental abilities, and environmental factors such as cultural milieu and availability of resources. A more general model—the *Comprehensive Model of Talent Development* (Gagné, 2015)—describes the progression from biological foundations to gifts and then to talents. Simonton (2014) has proposed a somewhat similar model to direct research on individual differences in creative performance. This model posits that both environmental and genetic factors impact creative performance through cognitive abilities and dispositional traits, which may impact performance directly and through deliberate practice.

Another important theoretical framework is Zielger and colleagues' *Actiotope Model of Giftedness* (AMG). The AMG posits that excellence in a domain reflects not only individual difference characteristics, but also the person's entire *actiotope,* which comprises the "actions and the possibilities for acting possessed by individuals" and "an individual and the material, social and informational environment in which that individual actively interacts" (Ziegler, Vialle, & Wimmer, 2013, p. 3). (The inspiration for the term *actiotope* is *biotope*, which is very similar in meaning to *habitat.*) Zielger and colleagues differentiate between two types of "capital" that impact learning—*learning capital* and *educational capital.* Learning capital includes characteristics of learners themselves—for example, a learner's goals (*telic capital,* from the Greek "telos," for goal or purpose) and their attentional resources (*attentional learning capital*). By contrast, educational capital includes characteristics of learners' environments—for example, investment in schools (*economic educational capital*) and the availability of institutions and individuals to support the learning process (*social educational capital*).

Several other models may be useful for identifying possible determinants of individual differences in expertise. One such model is Ackerman and colleagues' PPIK theory—*Intelligence-as-Process, Personality, Interests, and Intelligence-as-Knowledge* (see Ackerman, 1996; Ackerman & Beier, 2004). Extending Cattell's (1971) *Investment Theory of Intelligence,* PPIK posits that the acquisition of domain knowledge—the core factor of the adult intellect—is influenced by constellations of personality traits, abilities, and interests (or "trait complexes"). In a model of occupational and academic performance, Schmidt (2014) proposed that introversion and fluid intelligence are the primary causes of general interest in learning (intellectual curiosity). Intellectual curiosity is the primary cause of crystallized intelligence (including knowledge), which in turn is the primary cause of occupational and academic performance. Gardner's (1983, 1999) *Theory of Multiple Intelligences* could also be used as a guide for identifying basic abilities that are relevant to acquiring expertise in various domains, as could Sternberg's (1985) *Triarchic Theory of Intelligence,* which posits analytical, practical, and creative intelligences. Côté and colleagues' *Developmental Model of Sport Participation* could be used to identify forms of domain-relevant experience other than deliberate practice that may predict expertise (e.g., deliberate play; see Côté et al., 2007).

Task and situational factors must also be included in a comprehensive account of expertise. Feltovich and colleagues' analysis of situations that create difficulty in developing high levels of proficiency in medical practice is particularly relevant (Feltovich, Coulson, & Spiro, 2001; see also Hoffman et al., 2014). Through cognitive task analysis and interviews with subject matter experts, these researchers identified several "dimensions of difficulty" that make tasks require mental effort, such as *static versus dynamic, discrete versus continuous,* and *sequential versus simultaneous.* A key question for future research is whether these task/situational factors interact with individual-difference factors in predicting expertise. For example, as alluded to earlier, a static task environment may afford greater use of retrieval structures (i.e., Ericsson & Kintsch's, 1995, long-term working memory) to encode and maintain rapid access to domain-relevant information than a dynamic task environment, making circumvention of working memory limitations possible in the former situations but not the latter.

5.2 Multifactorial Gene–Environment Interaction Model

Figure 6 illustrates a model for investigating expertise from an integrative perspective. The model is an instantiation of the *Multifactorial Gene–Environment Interaction Model* (MGIM) of expertise recently proposed by Ullén, Hambrick, and Mosing (2015). The MGIM considers expertise at both the genotypic and phenotypic levels and assumes genetic factors, nongenetic factors, and their interactions to be important influences throughout the model. The MGIM includes both *distal* predictors of expertise—factors that influence performance indirectly through other factors (e.g., opportunity factors, deliberate practice)—and *proximal* predictors of performance—factors that may influence performance directly (e.g., basic ability factors, domain-specific knowledge factors). It also includes factors that are both proximal *and* distal predictors, in that they have both direct and indirect effects on performance (e.g., ability factors, personality factors). Finally, the model captures interactions between factors (e.g., task/situational × ability interactions).

As illustrated in the reanalysis of the results of Masunaga and Horn's (2001) study of GO expertise (see Figure 5), SEM is the optimal approach for testing the MGIM. Along with permitting analysis of data at the latent level, SEM permits analysis of either cross-sectional or longitudinal data. It also allows the researcher to test multiple relationships involving multiple variables in a single analysis and to statistically compare different configurations of variables representing competing models. Finally, SEM can be used to test for nonlinear relationships, such as interactions between individual difference and task/situational variables. In short, SEM can accommodate most any type of quantitative data (behavioral, genetic, or neural) collected in studies of expertise and can be used to test for many different types of relationships.

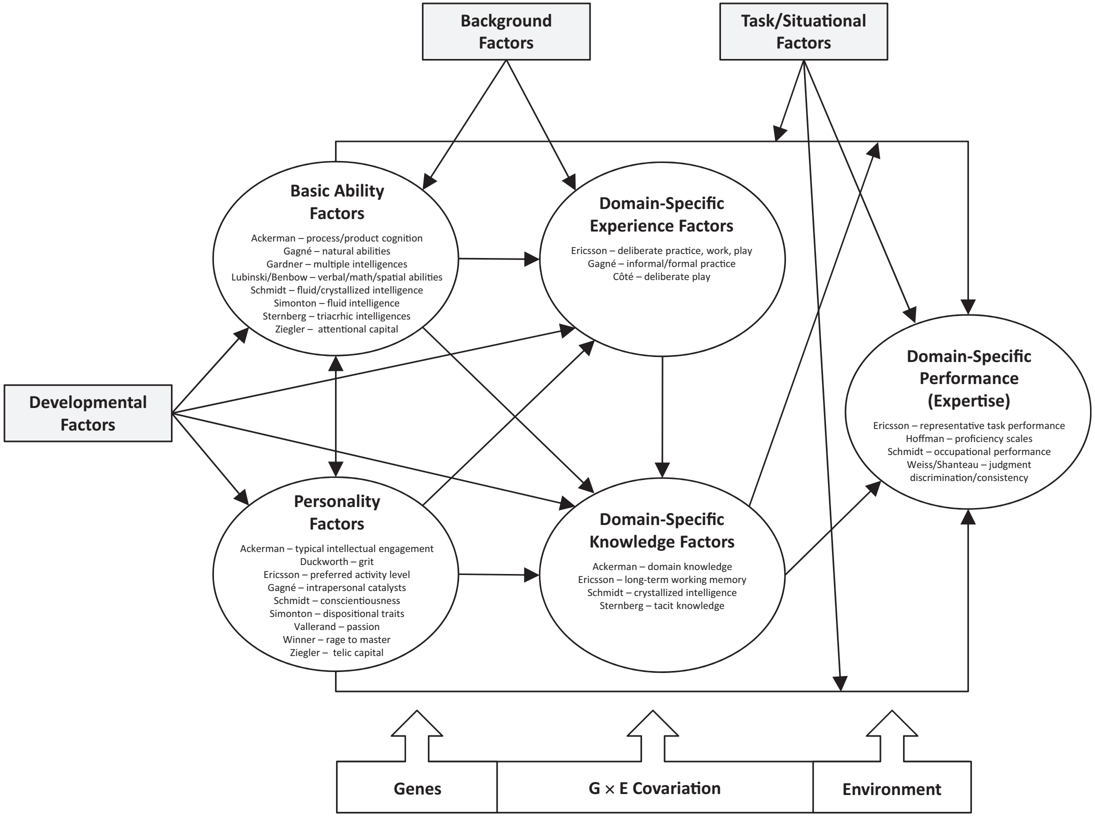

Figure 6 Instantiation of the multifactorial gene–environment interaction model (MGIM; Ullén et al., 2015) illustrating possible influences of different types of factors on expertise (names refer to associated theorists).

5.3 A Mathematical Simulation Approach

Mathematical simulation offers another way to investigate the origins of expertise. One of us (Campitelli, Gobet, & Bilalić, 2014) has used such an approach to investigate how well three models of chess expertise capture five major results found in the chess expertise literature: (1) there are very young chess players with remarkable achievements; (2) there is a moderate correlation between chess rating and both chess playing and chess practice; (3) there is a critical period (i.e., players that start practicing chess at earlier ages are much more likely to achieve higher levels of expertise in chess); (4) the correlation between chess expertise and intelligence is much higher in children than in adults; and (5) there is a decline in chess skill in older adults.

The simulation included three competing models. The *practice-motivation (PM) model* was the implementation of the deliberate practice view. The only differences among individuals in this model are the number of hours of practice and the motivation toward engaging in practice. The *practice-intelligence (PI) model* explains differences in chess skill based on differences in practice, playing, and intelligence. Finally, the *practice-plasticity-processes (PPP) model* indicates that the differences in chess skill are accounted for by differences in practice, playing, and neural plasticity. Differences in neural plasticity lead to differences in the acquisition of domain-specific patterns and heuristics. Note that in this model the correlation between chess skill and intelligence is due to the fact that neural plasticity affects both the acquisition of chess skill and intelligence.

The results of the simulation indicate that PPP was able to capture four of the five criteria (all but the fourth, which none of the models captured). By contrast, PM and PI captured at most three criteria. Obviously, a mathematical simulation is not a behavioral test of the models, but the simulation clearly shows that a model that only explains individual differences in chess skill based on differences in practice would never be able to account for well-established findings in the literature. This approach is interesting in its own right and could also be used to guide specification of factors and models in SEM research.

6. BEYOND EXPERTS ARE BORN VERSUS MADE

Our review of evidence concerning the origins of expertise can be summarized in the following points:

1. Though undeniably important from a statistical and practical perspective, deliberate practice does not account for all, nearly all, or even the majority of the variance in expertise.
2. Basic abilities predict expertise in some domains and sometimes even in highly skilled performers.
3. Personality factors predict expertise indirectly through deliberate practice, but may also predict expertise directly.
4. Forms of domain-relevant experience other than deliberate practice (e.g., work) positively and meaningfully predict expertise.
5. Genetically influenced factors account for individual differences in expertise, both indirectly through training and directly.
6. Models of expertise that only take into account deliberate practice will never adequately account for the major facts of expertise.

Over the past several decades, expertise has emerged as a vital area of psychological research, and all indications are that interest is only increasing, as evidenced by recent special issues of the journals *Intelligence* (Detterman, 2014) and *Frontiers in Psychology* (Campitelli, Connors, Bilalic, & Hambrick, 2014) devoted to the topic (see also Kaufman, 2013, for an edited volume). The intensity of the debate over the origins of expertise may to some degree reflect pre-theoretical biases—beliefs about why some people are more successful in various realms than others. No scientist can claim to be totally immune from such biases, and it is important to remain cognizant of such biases to maintain objectivity in this area of research. The debate may also reflect a clash between the two disciplines of scientific psychology—the *experimental* and the *differential* (Cronbach, 1957). Experimental psychologists seek to identify general trends in data, as reflected in means, whereas differential psychologists seek to explain individual differences, as reflected in measures of variance. Adherents of these approaches may "talk past each other."

Whatever the case, we believe that the most fruitful approach for future research on the origins of expertise is to embrace the idea that expertise is, at its core, a multiply determined phenomenon whose richness and complexity can never be adequately understood by focusing on one, or one class, of determinant, or by using one methodological approach. The experts are born versus made debate is over, and now the task for scientists is to develop and test theories that take into account the myriad ways that experts are born *and* made, using the most appropriate methodological approaches to test these theories. A respectful and open-minded exchange of ideas among researchers with diverse perspectives on expertise will make this a reality and advance understanding of topic of enduring fascination to scientists and non-scientists alike.

ACKNOWLEDGMENTS

We thank Lauren Harris and Fred Oswald for comments on an earlier version of this chapter.

REFERENCES

Ackerman, P. L. (1987). Individual differences in skill learning: an integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3–27. http://dx.doi.org/10.1037/0033-2909.102.1.3.

Ackerman, P. L. (1996). A theory of adult intellectual development: process, personality, interests, and knowledge. *Intelligence, 22*, 227–257. http://dx.doi.org/10.1016/S0160-2896(96)90016-1.

Ackerman, P. L. (2014). Facts are stubborn things. *Intelligence, 45*, 104–106. http://dx.doi.org/10.1016/j.intell.2014.01.002.

Ackerman, P. L., & Beier, M. E. (2004). Knowledge and intelligence. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 125–139). Thousand Oaks, CA: Sage.

Agdestein, S. (2013). *How Magnus Carlsen became the youngest chess grandmaster in the world. The story and the games*. Alkmaar, The Netherlands: New In Chess.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406. http://dx.doi.org/10.1037/0033-295X.89.4.369.

Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York: John Wiley and Sons.

Ando, J., Ono, Y., & Wright, M. J. (2001). Genetic structure of spatial and verbal working memory. *Behavior Genetics, 31*, 615–624. http://dx.doi.org/10.1023/A:1013353613591.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). Amsterdam: Elsevier.

Barnsley, R. H., Thompson, A. H., & Barnsley, P. E. (1985). Hockey success and birthdate: the relative age effect. *Canadian Association for Health, Physical Education, and Recreation Journal, 51*, 23–28.

Berri, D. J., & Simmons, R. (2011). Catching a draft: On the process of selecting quarterbacks in the National Football League amateur draft. *Journal of Productivity Analysis, 35*, 37–49. http://dx.doi.org/10.1007/s11123-009-0154-6.

Blokland, G. A. M., Mosing, M. A., Verweij, K. J. H., & Medland, S. (2013). Twin studies and behavior genetics. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 2, pp. 198–218). New York: Oxford University Press.

Bloom, B. S. (1985). Generalizations about talent development. In B. S. Bloom (Ed.), *Developing talent in young people*. New York: Ballantine Books.

Bonneville-Roussy, A., Lavigne, G. L., & Vallerand, R. J. (2011). When passion leads to excellence: the case of musicians. *Psychology of Music, 39*, 123–138. http://dx.doi.org/10.1177/0305735609352441.

Boot, W., & Ericsson, K. A. (2013). Expertise. In J. D. Lee, & A. Kirlik (Eds.), *The Oxford handbook of cognitive engineering* (pp. 143–158). http://dx.doi.org/10.1093/oxfordhb/9780199757183.013.0009.

Bouchard, T. J., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology, 54*, 4–45. http://dx.doi.org/10.1002/neu.10160.

de Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (2007). The influence of achievement motivation and chess-specific motivation on deliberate practice. *Journal of Sport and Exercise Psychology, 29*, 561–583.

de Bruin, A. B. H., Smits, N., Rikers, R. M. J. P., & Schmidt, H. G. (2008). Deliberate practice predicts performance over time in adolescent chess players and drop-outs: a linear mixed model analysis. *British Journal of Psychology, 99*, 473–497. http://dx.doi.org/10.1348/000712608X295631.

Campitelli, G., Connors, M. H., Bilalic, M., & Hambrick, D. Z. (2014). Psychological perspectives on expertise. *Frontiers in Psychology, 6*. http://dx.doi.org/10.3389/fpsyg.2015.00258.

Campitelli, G., & Gobet, F. (2011). Deliberate practice: necessary but not sufficient. *Current Directions in Psychological Science, 20*, 280–285. http://dx.doi.org/10.1177/0963721411421922.

Campitelli, G., Gobet, F., & Bilalić, M. (2014). Cognitive processes and development of chess genius: an integrative approach. In D. K. Simonton (Ed.), *The Wiley handbook of genius* (pp. 350–374). Chichester, UK: John Wiley & Sons. http://dx.doi.org/10.1002/9781118367377.ch17.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Oxford, England: Houghton Mifflin.

Charness, N., Krampe, R. Th, & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: an international comparison of life-span chess skill acquisition. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 51–80). New York: Psychology Press.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19*, 151–165. http://dx.doi.org/10.1002/acp.1106.

Chase, W. G., & Simon, H. A. (1973a). Perception in chess. *Cognitive Psychology, 4*, 55–81.

Chase, W. G., & Simon, H. A. (1973b). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 216–281). New York: Academic Press.

Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy, 52*, 337–345. http://dx.doi.org/10.1037/pst0000015.

Coon, H., & Carey, G. (1989). Genetic and environmental determinants of musical ability in twins. *Behavior Genetics, 19*, 183–193. http://dx.doi.org/10.1007/BF01065903.

Côté, J., Baker, J., & Abernethy, B. (2007). Practice and play in the development of sport expertise. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of sport psychology* (Vol. 3, pp. 184–202). Hoboken, NJ: John Wiley & Sons.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684. http://dx.doi.org/10.1037/h0043943.

Der Spiegel. (1987). Genieblitze und Blackouts: Der Spiegel testete Intelligenz, Gedächtnis, und Schachkunst Garri Kasparows (Strokes of genius and blackouts: Der Spiegel tests the intelligence, memory, and chess ability of Garry Kasparov). *Der Spiegel, 41*(1987), 126–140.

Detterman, D. K. (2014). Introduction to the intelligence special issue on the development of expertise: Is ability necessary? *Intelligence, 45*, 1–5. http://dx.doi.org/10.1016/j.intell.2014.02.004.

Detterman, D. K., Gabriel, L. T., & Ruthsatz, J. M. (1998). Absurd environmentalism. *Behavioral and Brain Sciences, 21*, 411–412. http://dx.doi.org/10.1017/S0140525X98271238.

Di Rosa, C., Cieri, F., Antonucci, I., Stuppia, L., & Gatta, V. (2015). Music in DNA: from Williams Syndrome to musical genes. *Open Journal of Genetics, 5*, 12. http://dx.doi.org/10.4236/ojgen.2015.51002.

Dodson, J. (2005). *Ben Hogan: An American life*. New York: Broadway Books.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science, 291*(5510), 1969–1972.

Duckworth, A. L., & Gross, J. J. (2014). Self-control and grit related but separable determinants of success. *Current Directions in Psychological Science, 23*, 319–325. http://dx.doi.org/10.1177/0963721414541462.

Duckworth, A. L., Kirby, T. A., Tsukayama, E., Berstein, H., & Ericsson, K. A. (2012). Deliberate practice spells success: why grittier competitors triumph at the National Spelling Bee. *Social Psychological and Personality Science, 2*, 174–181. http://dx.doi.org/10.1177/1948550610385872.

Duffy, L. J., Baluch, B., & Ericsson, K. A. (2004). Dart performance as a function of facets of practice amongst professional and amateur men and women players. *International Journal of Sport Psychology, 35*, 232–245.

Engle, R. W. (2002). WMC as executive attention. *Current Directions in Psychological Science, 11*, 19–23. http://dx.doi.org/10.1111/1467-8721.00160.

Epstein, D. (2014). *The sports gene: Inside the science of extraordinary athletic performance*. New York: Penguin Group.

Ericsson, K. A. (2005). Recent advances in expertise research: a commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19*, 233–241. http://dx.doi.org/10.1002/acp.1111.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). New York: Cambridge University Press.

Ericsson, K. A. (2007). Deliberate practice and the modifiability of body and mind: toward a science of the structure and acquisition of expert and elite performance. *International Journal of Sport Psychology, 38*, 4–34.

Ericsson, K. A. (1998). The scientific study of expert levels of performance: general implications for optimal learning and creativity. *High Ability Studies, 9*, 75–100. http://dx.doi.org/10.1080/1359813980090106.

Ericsson, K. A. (2012). *The danger of delegating education to journalists: Why the APS observer needs peer review when summarizing new scientific developments* (Unpublished manuscript.) Retrieved from: https://psy.fsu.edu/faculty/ericsson/ericsson.hp.html.

Ericsson, K. A. (2013a). My exploration of Gagné's "evidence" for innate talent: It is Gagné who is omitting troublesome information so as to present more convincing accusations. In S. B. Kaufman (Ed.), *Beyond talent or practice: The complexity of greatness* (pp. 223–256). New York: Oxford University Press.

Ericsson, K. A. (2013b). Training history, deliberate practice and elite sports performance: an analysis in response to Tucker and Collins review—what makes champions? *British Journal of Sport Medicine, 47*, 533–535. http://dx.doi.org/10.1136/bjsports-2012-091767.

Ericsson, K. A. (2014a). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: a response to criticisms. *Intelligence, 45*, 81–103. http://dx.doi.org/10.1016/j.intell.2013.12.001.

Ericsson, K. A. (2014b). *Challenges for the estimation of an upper-bound on relations between accumulated deliberate practice and the associated performance of novices and experts: Comments on Macnemara (sic), Hambrick, and Oswald's (2014) published meta-analysis* (Unpublished manuscript.) Retrieved from: https://psy.fsu.edu/faculty/ericsson/ericsson.hp.html.

Ericsson, K. A. (2014c). *Supplemental online materials for "A challenge to estimates of an upper-bound on relations between accumulated deliberate practice and the associated performance in domains of expertise: Comments on Macnemara (sic), Hambrick, and Oswald's (2014) published meta-analysis."* (Unpublished manuscript.) Retrieved from: https://psy.fsu.edu/faculty/ericsson/ericsson.hp.html.

Ericsson, K. A., & Charness, N. (1994). Expert performance: its structure and acquisition. *American Psychology, 49*, 725–747. http://dx.doi.org/10.1037/0003-066X.49.8.725.

Ericsson, K. A., & Charness, N. (1995). "Expert performance: its structure and acquisition": reply. *American Psychologist, 50*, 803–804.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245. http://dx.doi.org/10.1037/0033-295X.102.2.211.

Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406. http://dx.doi.org/10.1037/0033-295X.100.3.363.

Ericsson, K. A., & Moxley, J. H. (2012). A critique of Howard's argument for innate limits in chess performance or why we need an account based on acquired skill and deliberate practice. *Applied Cognitive Psychology, 26*, 649–653. http://dx.doi.org/10.1002/acp.2841.

Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2005). Giftedness viewed from the expert-performance perspective. *Journal for the Education of the Gifted, 28*, 287–311. http://dx.doi.org/10.4219/jeg-2005-335.

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard Business Review, 85*, 146–147.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Towne, T. J. (2010). Expertise. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 404–416. http://dx.doi.org/10.1002/wcs.47.

Feltovich, P. J., Coulson, R. L., & Spiro, R. J. (2001). Learners' (mis)understanding of important and difficult concepts: a challenge to smart machines in education. In K. D. Forbus, & P. J. Feltovich (Eds.), *Smart machines in education* (pp. 349–375). Cambridge, MA: MIT Press.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Oxford: Brooks/Cole.

Ford, P. R., & Williams, A. M. (2012). The developmental activities engaged in by elite youth soccer players who progressed to professional status compared to those who did not. *Psychology of Sport and Exercise, 13*, 349–352. http://dx.doi.org/10.1016/j.psychsport.2011.09.004.

Foroughi, C. K., Werner, N. E., Barragán, D., & Boehm-Davis, D. A. (2015). Interruptions disrupt reading comprehension. *Journal of Experimental Psychology: General, 144*, 704–709. http://dx.doi.org/10.1037/xge0000074.

Fox, P. W., Hershberger, S. L., & Bouchard, T. J. J. (1996). Genetic and environmental contributions to the acquisition of a motor skill. *Nature, 384*(6607), 356–358.

Gagné, F. (2007). Predictably, an unconvincing second attempt. *High Ability Studies, 18*, 67–69. http://dx.doi.org/10.1080/13598130701350742.

Gagné, F. (2009). Debating giftedness: pronat vs. antinat. In L. V. Shaninina (Ed.), *International handbook on giftedness* (pp. 155–198). Dordrecht, Netherlands: Springer.

Gagné, F. (2013). The DMGT: changes within, beneath, and beyond. *Talent Development and Excellence, 5*, 5–19.

Gagné, F. (2014). Yes, giftedness (aka "innate" talent) does exist! In S. B. Kaufman (Ed.), *Beyond talent or practice: The complexity of greatness* (pp. 191–222). New York: Oxford University Press.

Gagné, F. (2015). From genes to talent: the DMGT/CMTD perspective. Critical issues on gifted education and talent development. *Revista de Educación, 368*, 12–37. http://dx.doi.org/10.4438/1988-592X-RE-2015-368-289.

Galton, F. (1869). *Hereditary genius*. London: Macmillan.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gardner, H. (1995). "Expert performance: its structure and acquisition": comment. *American Psychologist, 50*, 802–803. http://dx.doi.org/10.1037/0003-066X.50.9.802.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st Century*. New York: Basic Books.

Gladwell, M. (2008). *Outliers: The story of success*. New York: Little, Brown, and Co.

Gobet, F., & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology, 43*, 159–172. http://dx.doi.org/10.1037/0012-1649.43.1.159.

Gobet, F., & Ereku, M. H. (2014). Checkmate to deliberate practice: the case of Magnus Carlsen. *Frontiers in Psychology, 5*, 878. http://dx.doi.org/10.3389/fpsyq.2014.00878.

Goldsmith, B. (2005). *Obsessive genius: The inner world of Marie Curie (great discoveries)*. New York: W.W. Norton & Company.

Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: a psychometric investigation. *Acta Psychologica, 124*, 398–420. http://dx.doi.org/10.1016/j.actpsy.2006.07.008.

de Groot, A. D. (1946/1978). *Thought and choice in chess*. Berlin, Germany: De Gruyter Mouton.

Halam, S. (1998). The predictors of achievement and dropout in instrumental tuition. *Psychology of Music, 26*, 116–132. http://dx.doi.org/10.1177/0305735698262002.

Hambrick, D. Z., Altmann, E. M., Oswald, F. L., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Accounting for expert performance: the devil is in the details. *Intelligence, 45*, 112–114. http://dx.doi.org/10.1016/j.intell.2014.01.007.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, WMC, and age on cognitive performance: an investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*, 339–387. http://dx.doi.org/10.1006/cogp.2001.0769.

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). London: Cambridge Press.

Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., et al. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: the case of geological bedrock mapping. *Journal of Experimental Psychology: General, 141*, 397–403. http://dx.doi.org/10.1037/a0025927.

Hambrick, D. Z., & Macnamara, B. N. (2016). Expertise. In S. K. Whitbourne (Ed.), *The encyclopedia of adulthood and aging* (Vol. 3, pp. 466–471). Oxford, UK: John Wiley and Sons.

Hambrick, D. Z., & Meinz, E. J. (November 20, 2011a). Sorry, strivers. Talent matters. *The New York Times*. Sunday Review, 12.

Hambrick, D. Z., & Meinz, E. J. (2011b). Limits of the predictive power of domain-specific experience. *Current Directions in Psychological Science, 20*, 275–279. http://dx.doi.org/10.1177/0963721411422061.

Hambrick, D. Z., & Meinz, E. J. (2012). Working memory capacity and musical skill. In T. P. Alloway, & R. G. Alloway (Eds.), *Working memory: The connected intelligence* (pp. 137–155). New York: Psychology Press.

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence, 45*, 34–45. http://dx.doi.org/10.1016/j.intell.2013.04.001.

Hambrick, D. Z., & Tucker-Drob, E. (2014). The genetics of music accomplishment: evidence for gene-environment correlation and interaction. *Psychonomic Bulletin and Review*. http://dx.doi.org/10.3758/s13423-014-0671-9.

Harris, J. A., Vernon, P. A., Johnson, A. M., & Jang, K. L. (2006). Phenotypic and genetic relationships between vocational interests and personality. *Personality and Individual Differences, 40*, 1531–1541. http://dx.doi.org/10.1016/j.paid.2005.11.024.

Haugaasen, M., Toering, T., & Jordet, G. (2014). From childhood to senior professional football: a multi-level approach to elite youth football players' engagement in football-specific activities. *Psychology of Sport and Exercise, 15*, 336–344. http://dx.doi.org/10.1016/j.psychsport.2014.02.007.

Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. H. (2014). *Accelerated expertise: Training for high proficiency in a complex world*. New York: Psychology Press.

Howard, R. W. (2011). Does high-level intellectual performance depend on practice alone? Debunking the Polgár sisters case. *Cognitive Development, 26*, 196–202. http://dx.doi.org/10.1016/j.cogdev.2011.04.001.

Howard, R. W. (2012). Longitudinal effects of different types of practice on the development of chess expertise. *Applied Cognitive Psychology, 26*, 359–369. http://dx.doi.org/10.1002/acp.1834.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park: Sage.

Johnson, M. B., Tenenbaum, G., & Edmonds, W. A. (2006). Adaptation to physically and emotionally demanding conditions: the role of deliberate practice. *High Ability Studies, 17*, 117–136. http://dx.doi.org/10.1080/13598130600947184.

Kaufman, S. B. (2013). *The complexity of greatness: Beyond talent or practice*. New York: Oxford University Press.

Kaufman, S. B. (2014). A proposed integration of the expert performance and individual differences approaches to the study of elite performance. *Frontiers in Psychology, 5*, 707. http://dx.doi.org/10.3389/fpsyq.2014.00707.

Keith, N., & Ericsson, K. A. (2007). A deliberate practice account of typing proficiency in everyday typists. *Journal of Experimental Psychology: Applied, 13*, 135–145. http://dx.doi.org/10.1037/1076-898X.13.3.135.

Klein, A. M. (1993). *Sugarball: The American game, the Dominican dream*. New Haven, CT: Yale University Press.

Kline, R. B. (2011). *Structural equation modeling: Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Kopiez, R., & Lee, J. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research, 8*, 97–120. http://dx.doi.org/10.1080/14613800600570785.

Krampe, R. Th. (1994). *Maintaining excellence: Cognitive-motor performance in pianists differing in age and skill level*. Berlin: Max-Planck-Institut für Bildungsforschung. Edition Sigma. Retrieved from: http://pubman.mpdl.mpg.de/pubman/item/escidoc:2103276/component/escidoc:2103275/Studien_Berichte_MPIB_058.pdf.

Krampe, R. Th., & Ericsson, K. A. (1996). Maintaining excellence: deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General, 125*, 331–359. http://dx.doi.org/10.1037/0096-3445.125.4.331.

Kremen, W. S., Jacobsen, K. C., Zian, H., Eisen, S. A., Eaves, J. J., Tsuang, M. T., et al. (2007). Genetics of verbal working memory processes: a twin study of middle-aged men. *Neuropsychology, 21*, 569–580. http://dx.doi.org/10.1037/0894-4105.21.5.569.

Law, M. P., Côté, J., & Ericsson, K. A. (2007). Characteristics of expert development in rhythmic Gymnasts: a retrospective study. *International Journal of Sport and Exercise Psychology, 5*, 82–103. http://dx.doi.org/10.1080/1612197X.2008.9671814.

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, environment, and personality: A study of 850 sets of twins*. Austin: University of Texas Press.

Lombardo, M. P., & Deaner, R. O. (2014). You can't teach speed: sprinters falsify the deliberate practice model of expertise. *PeerJ, 2*, e445. http://dx.doi.org/10.7717/peerj.445.

Lubinski, D. (2009). Exceptional cognitive ability: the phenotype. *Behavior Genetics, 39*, 350–358. http://dx.doi.org/10.1007/s10519-009-9273-0.

Lubinski, D., Benbow, C. P., & Kell, H. J. (2014). Life paths and accomplishments of mathematically precocious males and females four decades later. *Psychological Science, 25*, 2217–2232. http://dx.doi.org/10.1177/0956797614551371.

Lyons, B., Hoffman, B., & Michel, J. (2009). Not much more than g? An examination of the impact of intelligence on NFL performance. *Human Performance, 22*, 225–245. http://dx.doi.org/10.1080/08959280902970401.

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychological Science, 25*, 1608–1618. http://dx.doi.org/10.1177/0956797614535810.

Macnamara, B. N., Moreau, D., & Hambrick, D. Z. (2015). The relationship between deliberate practice and performance in sports: a meta-analysis. *Perspectives on Psychological Science* (in press).

Marcus, G. (2012). *Guitar zero: The science of becoming musical at any age*. New York: Penguin.

Masunaga, H., & Horn, J. (2001). Expertise and age-related changes in components of intelligence. *Psychology and Aging, 16*, 293–311. http://dx.doi.org/10.1037/0882-7974.16.2.293.

McLaughlin, D. (2014). *The first half of a journey in human potential: Half way to the 10,000 hour goal, four years of a blog by Dan McLaughlin (the Dan plan)* (Kindle). Retrieved from: http://www.amazon.com/First-Half-Journey-Human-Potential-ebook/dp/B00MTC0NJA.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: the role of WMC. *Psychological Science, 21*, 914–919. http://dx.doi.org/10.1177/0956797610373933.

Meinz, E. J., Hambrick, D. Z., Hawkins, C. B., Gillings, A. K., Meyer, B. E., & Schneider, J. L. (2012). Roles of domain knowledge and WMC in components of skill in Texas hold'em poker. *Journal of Applied Research in Memory and Cognition, 1*, 34–40. http://dx.doi.org/10.1016/j.jarmac.2011.11.001.

Mendel, G. (1866). *Versuche über Plflanzenhybriden: Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865* (pp. 3–47). Abhandlungen http://www.esp.org/foundations/genetics/classical/gm-65.pdf.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: no causal effect of music practice on music ability. *Psychological Science, 25*, 1795–1803. http://dx.doi.org/10.1177/0956797614541990.

Moxley, J. H., & Charness, N. (2013). Meta-analysis of age and skill effects on recalling chess positions and selecting the best move. *Psychonomic Bulletin and Review, 20*, 1017–1022. http://dx.doi.org/10.3758/s13423-013-0420-5.

Moxley, J. H., Ericsson, K. A., Scheiner, A., & Tuffiash, M. (2015). The effects of experience and disuse on crossword solving. *Applied Cognitive Psychology, 29*, 73–80. http://dx.doi.org/10.1002/acp.3075.

Neale, M. C., & Cardon. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Norman, G., & Phillips, D. T. (2006). *The way of the shark: Lessons on golf, business, and life*. New York: Atria Books.

Peltopuro, M., Ahonen, T., Kaartinen, J., Seppälä, H., & Närhi, V. (2014). Borderline intellectual functioning: a systematic literature review. *Intellectual and Developmental Disabilities, 52*, 419–443. http://dx.doi.org/10.1352/1934-9556-52.6.419.

Platz, F., Kopiez, R., Lehmann, A. C., & Wolf, A. (2014). The influence of deliberate practice on musical achievement: a meta-analysis. *Frontiers in Psychology, 5*, 646–658. http://dx.doi.org/10.3389/fpsyg.2014.00646.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral genetics* (5th ed.). New York: Worth Publishers.

Plomin, R., Shakeshaft, N. G., McMillan, A., & Trzaskowski, M. (2014). Nature, nurture, and expertise. *Intelligence, 45*, 46–59. http://dx.doi.org/10.1016/j.intell.2013.06.008.

Plomin, R., & Spinath, F. M. (2004). Intelligence: genetics, genes, and genomics. *Journal of Personality and Social Psychology, 86*, 112–129. http://dx.doi.org/10.1037/0022-3514.86.1.112.

Polderman, T. J., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., et al. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics, 47*, 702–709. http://dx.doi.org/10.1038/ng.3285.

Polderman, T. J., Stins, J. F., Posthuma, D., Gosso, M. F., Verhulst, F. C., & Boomsma, D. I. (2006). The phenotypic and genotypic relation between working memory speed and capacity. *Intelligence, 34*, 549–560. http://dx.doi.org/10.1016/j.intell.2006.03.010.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods, 10*, 178–192. http://dx.doi.org/10.1037/1082-989X.10.2.178.

Riach, S. (2003). *Life lessons from the game of golf*. Colorado Springs: David C. Cook.

Robbins, T. W., Henderson, E. J., Barker, D. R., Bradley, A. C., Fearneyhough, C., Henson, R., et al. (1996). Working memory in chess. *Memory and Cognition, 24*, 83–93.

Robertson, K. F., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis: even among gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science, 19*, 346–351. http://dx.doi.org/10.1177/0963721410391442.

Roring, R. W., & Charness, N. (2007). A multilevel model analysis of expertise in chess across the life span. *Psychology and Aging, 22*, 291–299. http://dx.doi.org/10.1037/0882-7974.22.2.291.

Ruthsatz, J., & Detterman, D. K. (2003). An extraordinary memory: the case study of a musical prodigy. *Intelligence, 31*, 509–518. http://dx.doi.org/10.1016/S0160-2896(03)00050-3.

Ruthsatz, J., Ruthsatz-Stephens, K., & Ruthsatz, K. (2014). The cognitive bases of exceptional abilities in child prodigies by domain: similarities and differences. *Intelligence, 44*, 11–14.

Ruthsatz, J., & Urbach, J. B. (2012). Child prodigy: a novel cognitive profile places elevated general intelligence, exceptional working memory and attention to detail at the root of prodigiousness. *Intelligence, 40*, 419–426. http://dx.doi.org/10.1016/j.intell.2012.06.002.

Santrock, J. W. (2012). *Lifespan development*. Boston: McGraw Hill Education.

Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: a commentary on four recent articles. *Perspectives on Psychological Science, 9*, 211–218. http://dx.doi.org/10.1177/1745691613518074.

Schmidt, F. L., & Hunter, J. (1999). Theory testing and measurement error. *Intelligence, 27*, 183–198. http://dx.doi.org/10.1016/S0160-2896(99)00024-0.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162–173. http://dx.doi.org/10.1037/0022-3514.86.1.162.

Schneider, W. (1997). The impact of expertise on performance: illustrations from developmental research on memory and sports. *High Ability Studies, 8*, 7–18. http://dx.doi.org/10.1080/1359813970080102.

Schneider, W. (1998). Innate talent or deliberate practice as determinants of exceptional performance: Are we asking the right question? *Behavioral and Brain Sciences, 21*, 423–424. http://dx.doi.org/10.1017/S0140525X98411233.

Schneider, W. (2015). *Memory development from early childhood through emerging adulthood*. New York: Springer.

Schneider, K., Bös, K., & Rieder, H. (1993). Performance prediction in young top athletes. In J. Beckmann, H. Strang, & E. Hahn (Eds.), *Aufmerksamkeit und Energetisierung. Facetten con Konzentration und Leistung* (pp. 277–299). Göttingen: Hogrefe.

Schulz, R., & Curnow, C. (1988). Peak performance and age among superathletes: track and field, swimming, baseball, tennis, and golf. *Journal of Gerontology, 43*, P113–P120. http://dx.doi.org/10.1093/geronj/43.5.P113.

Simonton, D. K. (1991). Career landmarks in science: Individual differences and interdisciplinary contrasts. *Developmental Psychology, 27*, 119–130. http://dx.doi.org/10.1037/0012-1649.27.1.119.
Simonton, D. K. (1999). Talent and its development: an emergenic and epigenetic model. *Psychological Review, 106*, 435–457. http://dx.doi.org/10.1037/0033-295X.106.3.435.
Simonton, D. K. (2014). Creative performance, expertise acquisition, individual differences, and developmental antecedents: an integrative research agenda. *Intelligence, 45*, 66–73. http://dx.doi.org/10.1016/j.intell.2013.04.007.
Sloboda, J. A. (1996). The acquisition of musical performance expertise: deconstructing the "talent" account of individual differences in musical expressivity. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and game* (pp. 107–126). Mahwah, NJ: Lawrence Erlbaum. Bottom of Form.
Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of practice in the development of performing musicians. *British Journal of Psychology, 87*, 287–309.
Soberlak, P., & Côté, J. (2003). The developmental activities of elite ice hockey players. *Journal of Applied Sport Psychology, 15*, 41–49. http://dx.doi.org/10.1080/10413200390180053.
Sohn, Y. W., & Doane, S. M. (2004). Memory processes of flight situation awareness: interactive roles of WMC, long-term working memory, and expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*, 461–475. http://dx.doi.org/10.1518/hfes.46.3.461.50392.
Sonnentag, S., & Kleine, B. M. (2000). Deliberate practice at work: a study with insurance agents. *Journal of Occupational and Organizational Psychology, 73*, 87–102. http://dx.doi.org/10.1348/096317900166895.
Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.
Sternberg, R. J. (1996). Costs of expertise. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 347–354). Mahwah, NJ: Lawrence Erlbaum.
Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education A proposed direction forward based on psychological science. *Psychological Science in the Public Interest, 12*, 3–54. http://dx.doi.org/10.1177/1529100611418056.
Szalavitz, M. (May 2013). *10,000 hours may not make a master after all*. Time. Retrieved from: http://healthland.time.com/2013/05/20/10000-hours-may-not-make-a-master-after-all/.
Takagi, H. (1997). *Cognitive aging: Expertise and fluid intelligence*. Los Angeles, CA: University of Southern California (Unpublished doctoral dissertation).
Tan, T. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology, 5*, 658. http://dx.doi.org/10.3389/fpsyg.2014.00658.
Toma, M., Halpern, D. F., & Berger, D. E. (2014). Cognitive abilities of elite nationally ranked scrabble and crossword experts. *Applied Cognitive Psychology, 28*, 727–737. http://dx.doi.org/10.1002/acp.3059.
Tucker, R., & Collins, M. (2012). What makes champions? A review of the relative contribution of genes and training to sporting success. *British Journal of Sports Medicine, 46*, 555–561. http://dx.doi.org/10.1136/bjsports-2011-090548.
Tuffiash, M., Roring, R. W., & Ericsson, K. A. (2007). Expert performance in SCRABBLE: implications for the study of the structure and acquisition of complex skills. *Journal of Experimental Psychology: Applied, 13*, 124–134. http://dx.doi.org/10.1037/1076-898X.13.3.124.
Ullén, F., Mosing, M. A., Holm, L., Eriksson, H., & Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish Musical

Discrimination Test. *Personality and Individual Differences, 63*, 87–93. http://dx.doi.org/10.1016/j.paid.2014.01.057.

Ullén, F., Hambrick, D. Z., & Mosing, M. A. (2015). Rethinking expertise: A multifactorial gene-environment interaction model of expert performance. *Psychological Bulletin*. [Epub ahead of print]. http://dx.doi.org/10.1037/bul0000033.

Ullén, F., Mosing, M. A., & Madison, G. (2015). Associations between motor timing, music practice, and intelligence studied in a large sample of twins. *Annals of the New York Academy of Sciences, 1337*, 125–129. http://dx.doi.org/10.1111/nyas.12630.

Unterrainer, J. M., Kaller, C. P., Halsband, U., & Rahm, B. (2006). Planning abilities and chess: a comparison of chess and non-chess players on the Tower of London task. *British Journal of Psychology, 97*, 299–311. http://dx.doi.org/10.1348/000712605X71407.

Unterrainer, J. M., Kaller, C. P., Leonhart, R., & Rahm, B. (2011). Revising superior planning performance in chess players: the impact of time restriction and motivation aspects. *American Journal of Psychology, 124*, 213–225. http://dx.doi.org/10.5406/amerjpsyc.124.2.0213.

Vallerand, R. J. (2015). *The psychology of passion: A dualistic model*. New York: Oxford University Press.

Van Natta, D. (2011). *Wonder girl: The magnificent sporting life of Babe Didrikson Zaharias*. New York: Little, Brown and Company.

Vinkhuyzen, A. E., van der Sluis, S., Posthuma, D., & Boomsma, D. I. (2009). The heritability of aptitude and exceptional talent across different domains in adolescents and young adults. *Behavior Genetics, 39*, 380–392. http://dx.doi.org/10.1007/s10519-009-9260-5.

Wai, J. (2014). What does it mean to be an expert? *Intelligence, 45*, 122–123.

Watson, J. B. (1930). *Behaviorism*. Chicago, IL: The University of Chicago Press.

Weiss, D. J., & Shanteau, J. (2014). Who's the best? A relativistic view of expertise. *Applied Cognitive Psychology, 28*, 447–457. http://dx.doi.org/10.1002/acp.3015.

Winner, E. (1996a). The rage to master: the decisive case for talent in the visual arts. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 271–301). Hillsdale, NJ: Erlbaum.

Winner, E. (1996b). *Gifted children*. New York: Basic Books.

Winner, E. (2000). The origins and ends of giftedness. *American Psychologist, 55*, 159–169. http://dx.doi.org/10.1037/0003-066X.55.1.159.

Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Easteal, S., et al. (2003). ACTN3 genotype is associated with human elite athletic performance. *The American Journal of Human Genetics, 73*, 627–631. http://dx.doi.org/10.1086/377590.

Yokom, G. (May 2008). *My shot: Larry Nelson*. Golf Digest. Retrieved from: http://www.golfdigest.com/magazine/2008-05/myshot_nelson.

Zaharias, B. D. (1955). *This life I've led: My autobiography*. Central Valley, CA: Oak Tree.

Ziegler, A., Vialle, W., & Wimmer, B. (2013). The actiotope model of giftedness: a short introduction to some central theoretical assumptions. In S. N. Phillipson, H. Stoeger, & A. Ziegler (Eds.), *Exceptionality in East Asia* (pp. 1–17). London: Routledge.

This page intentionally left blank

CHAPTER TWO

Explaining the Basic-Level Concept Advantage in Infants…or Is It the Superordinate-Level Advantage?

Gregory L. Murphy[1]
Department of Psychology, New York University, New York, NY, USA
[1]Corresponding author: E-mail: gregory.murphy@nyu.edu

Contents

Psychology of Learning and Motivation, Volume 64
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2015.09.002

Abstract

Much research in the 1970s and 1980s established that both children and adults prefer to classify objects at an intermediate or *basic level* of categorization such as dog or car. In adults, this leads to a performance advantage that has been found in dozens of experiments. However, research on infant concepts in the 1990s led to a very different consensus, that infants' first concepts—or at least some of their very early concepts—are higher-level or *superordinate* concepts, such as animal or vehicle. Surprisingly, this contradiction has not received very much attention, and both fields continue to report their conclusions as if this does not create a major problem for the theory of concepts. This chapter reviews the evidence for both claims and concludes that (1) the evidence for the basic-level advantage for children and adults is overwhelming and (2) the best way to resolve this conflict is to reinterpret the results of the infant studies. I argue that limitations of the paradigms available to study infants have systematically reduced the richness of the materials in a way that has made superordinates seem to be more readily learnable than they are in real life. This conclusion provides a more convincing story for how concepts develop over the life span.

1. INTRODUCTION

The human conceptual system is a powerful means of representing and generalizing knowledge. One important aspect of that system is the hierarchical organization into *taxonomies* that characterizes an important subset of our knowledge. As shown in Figure 1, some concepts can be organized

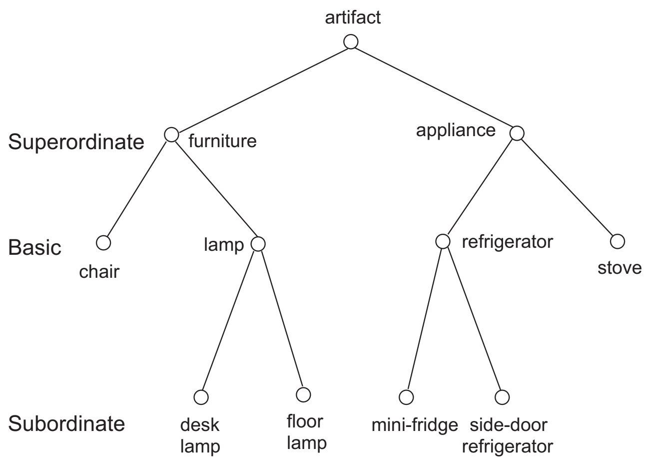

Figure 1 A fragment of the taxonomic organization of artifacts. The hierarchical structure allows inference, for example, if all appliances use electricity, then one can infer that side-door refrigerators use electricity.

into a treelike structure along a dimension of inclusiveness. For example, all desk lamps are lamps, all lamps are furniture, and all furniture is human artifacts. Not all concepts have inclusion relations. Furniture may also be works of art or school supplies, but not all furniture is a work of art, by any means, nor is all art furniture.

Among the categories that do fit into taxonomies are common categories of animals and objects that we encounter every day. Social or event categories may form (often shorter) taxonomies (e.g., holiday parties, parties, social events). In one of the most cited papers in cognitive psychology, Eleanor Rosch and her collaborators (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) proposed that one level of categorization in taxonomies is the primary one that we use to identify and think about the everyday objects in our world. This *basic level of categorization* consists of the most obvious and neutral way of identifying objects: chair, dog, boat, party, doctor, etc. Perhaps surprisingly, the basic level is not at a very high level of generality, like animal, plant, or machine, which they called *superordinate* categories. Neither is this neutral level the most specific one that we commonly form, with categories such as Scottish terrier, salad fork, or orthopedic surgeon, which they called the *subordinate level*.

Rosch et al. (1976) and many researchers afterward investigated the properties of the preferred level and how it differed from the other levels. In general, results showed that people preferred to use basic-level categories, that they were faster in doing so, that they learned basic categories more easily, and that category level had implications for language structure and use. These discoveries are now incorporated into the theory of human concepts (e.g., Mervis & Rosch, 1981; Murphy, 2002). As will be described in detail below, many developmental studies also investigated children's acquisition and use of concepts at the different levels. Very similar and often stronger results were found for school-age or preschool children: Children often knew and could use basic-level categories but did not know or had great difficulty with more general or specific concepts.

One discordant note in this research enterprise has been the claim that infants and young children form more general concepts first. These superordinate concepts are often called *global* concepts to indicate that they may not be identical to the adult superordinate concepts that they approximate. Although there is no strong consensus on just what those earliest concepts are, research has provided support for such global categories as land animals, sea animals, mammals, vehicles, furniture, or animals. Specific researchers have differed on exactly when they believe the global and basic categories appear. Jean Mandler (2004) proposes that true concepts begin around 1 year of life and tend to be global categories such as animals and furniture. Researchers on infant categorization propose that global categories can be acquired (under the right conditions) at 2 months of age, with basic categories following shortly afterward (e.g., Quinn & Johnson, 2000). Although these proposals are different in important respects, they are similar in that they argue that global categories are acquired first, in direct contradiction to research with preschoolers, which claims that children can learn and use basic categories with ease but superordinates only with great effort, if at all.

I believe that this is a real contradiction. It is not a result of researchers using terms in different ways—they are making truly incompatible arguments about conceptual development. And, what is worse, both have strong evidence for their claims. It is a bit surprising that this contradiction has not received more attention, as it suggests a serious problem in the theory of children's concepts, and, I believe, for adult concepts as well. There are very strong theoretical reasons and empirical evidence for believing in a basic-level advantage. If in fact superordinate-like concepts are the first to be learned in infancy, then those theoretical assumptions must be questioned, and the empirical evidence must be reconsidered. This is a serious problem, because the amount of evidence for the basic-level advantage is not small. It is a concern that such evidence could all be called into question, and it is also unclear how to explain how an immature child can easily do things that older children and adults find difficult to do.

The reason that this contradiction has not drawn more attention probably has to do with the division of scientific labor in which different researchers focus on different age groups. Infancy researchers do not have to worry about explaining the results of category learning in first graders. Researchers on language acquisition do not spend their time trying to explain dishabituation studies of 3-month-olds. The main research on basic-level categories in adults was carried out prior to the publication of the infant data. Thus, the contradiction exists across different laboratories and literatures rather than being present within the work of individual investigators, where it likely would have drawn more attention.

Most of the discussion of this problem has taken the form of criticism of specific findings, but the field has lacked a more general attempt to understand and attempt to resolve the broad findings that have created this disturbing contradiction. The present chapter attempts to do just that.

1.1 Goals of This Article

There is a gap in the theory of conceptual development that cannot be easily papered over. Researchers in adult concepts and semantic memory repeat or mutely assume the superiority of basic-level concepts in their work. This assumption infiltrates other topics in psychology, such as memory or visual search studies where researchers use basic-level names to ensure that subjects will easily identify their stimuli. Researchers in most of conceptual development find that basic-level concepts are easier to learn and process. Indeed, some have devoted years of their careers investigating how children manage to learn the other, more difficult concepts. However, there is evidence that basic-level categories are not the easiest to learn in infancy.

Perhaps there is an unusual developmental progression, in which superordinate categories are easy for infants to learn but after 2 years become difficult for some reason, with children responding at chance in many tasks and adults having significantly worse performance on them in almost every task. So far as I know, no one has explicitly proposed such a progression. To the degree that there is any acknowledgment of the strange change of the superordinate level's status within developmental psychology, it often takes the form of criticizing the notion of the basic level (e.g., Mandler, 2004).

This contradiction cannot be resolved by attempting to overturn the primary evidence, I will argue. The evidence is too strong. Instead, we will need to change some aspects of how we interpret that evidence. In particular, we will have to think carefully about what each kind of experiment tells us about what concepts people actually have and use. People may be able to do something in the lab but may not in fact do it under the conditions of everyday life. We will also have to critically examine the relation between the structure of experimental categories and those of the world outside the lab. Finally, we will have to face the question of what concepts are and what they are for. Although such foundational debates often cause more heat than light, I believe that this is necessary in order for us to be able to bring some light out of the present darkness.

I begin the chapter with a selective review of the two developmental literatures most relevant to the debate: initial findings of basic-level superiority in children and the infant category-learning literature. In general, I highlight only methodological issues that are relevant to our investigation and do not pretend to be providing a general review of these topics. An overview of both areas can be found in Murphy (2002). Then I more briefly and selectively discuss computational models that have found that superordinates may be learned first. Following these reviews, I provide an interpretation of the different literatures that attempts to provide a consistent story about children's cognitive capacities and how they lead to different concepts. The main merit I can claim for my proposal is that it exists, and so far as I know, there is no other attempt to encompass these two literatures into a single, coherent story that takes seriously the results in both.

1.2 The Nature of Hierarchically Organized Concepts

The question of whether global or basic concepts are primary takes place within a context of hierarchically structured categories like the ones shown in Figure 1. These are generally rich, family resemblance categories. There is a trivial sense in which any group of items that has any commonality or theme can be called a category. For example, all pink things form a category from this perspective, as do all things that are 937 cm tall, all rounded objects, all events that lasted longer than an hour, and so on. However, these are not usually what we consider to be psychological categories, because they are based on a single property that does not allow for further induction (Lassaline & Murphy, 1996; Murphy, 2010). Once you know that an object is a pink thing, you don't know anything further about it. In contrast, if you know an object is a shirt, you know its function, its likely size, what materials it is probably made out of, its shape, and so on. This inductive richness is present in most family-resemblance concepts. Of course, categorizing things as pink may be useful for a photo shoot, but pinkness is still not the concept we would use to identify the objects. For example, we would be more likely to say "Hand me that book over there" than "Hand me that pink thing over there."

Thus, when we look at the infant literature in particular, we will be focusing on evidence that babies learn rich categories like animals, dogs, shirts, or clothing and not on whether they can learn property-based categories like pink things.

2. DEVELOPMENTAL CATEGORY-LEVEL DIFFERENCES

Not that long ago, young children were not thought to have sophisticated cognitive capacities. The Piagetian program, in particular, argued that true concepts were not formed until quite late in childhood, and even those did not reach a sophisticated level until adolescence, when children could reason correctly about the logical relations among taxonomic categories (Inhelder & Piaget, 1964).

It was in this context that Rosch et al. (1976) pointed out that the failure of children to form concepts in the prior literature had been found almost exclusively with abstract properties (e.g., geometric blocks) or superordinate categories like animals or vehicles. They pointed out that researchers probably did not consider grouping together several shoes or several dogs as being a true categorization task, since the objects are so obviously the same kind of thing. Therefore Rosch et al. repeated some of those classic tasks using basic categories.

In Rosch et al.'s (1976) Experiment 8, 3- and 4-year-olds performed an oddity task in which two items were from the same category and the other from a different category. The two similar items were either from the same basic category (e.g., two fish) or the same superordinate (e.g., two animals). (See Murphy, 2002, p. 328, for a discussion of their design.) The 4-year-olds performed at ceiling, and the 3-year-olds were also at ceiling for basic trials but got only 55% of the superordinate trials correct. The animal—vehicle distinction tested was also used in a number of infant studies of global categories (see below).

In Experiment 9, Rosch et al. (1976) repeated the traditional sorting task in which children are given a set of pictures and have to group them into things that "are the same kind of thing." For basic-level categories, all but one kindergartner and first grader sorted correctly. For superordinates, only half of the children at each age sorted correctly.

Mervis and Crisafi (1982) followed up these results by testing children on novel stimuli and categories, which they constructed to be parallel to basic and superordinate categories. Children were shown one stimulus as a target and then asked which of the two other stimuli was the same kind of thing. They were able to identify the correct item when it was in the basic category virtually flawlessly. However, identifying an item in the same superordinate was more difficult, as 2-year-olds were accurate only 66% of the time and 4-year-olds 92% of the time. Horton and Markman (1980) found similar results in their sorting task with artificial stimuli.

Callanan and Markman (1982) asked children simple yes/no questions about familiar objects such as toys, cars, and animals, e.g., "Is this a doll?"; "Are these toys?" They discovered that children had difficulty with the superordinate questions about single objects. Blewitt (1994) also found such effects, with 2-year-olds responding at 92% and 64% correct to basic and superordinate categories chosen to be familiar, with similar results in later experiments. In a second study, Callanan and Markman asked children whether a puppet's sentences were OK or silly. Children were more accurate at identifying the sentences about basic-level categorization as OK than those about superordinate categorization.

Callanan (1989) labeled a familiar object with a novel name in puppet language and then asked children to decide whether other objects could also be called by the same name. Preschoolers predominantly interpreted this name as referring to subordinate or basic-level categories. Golinkoff, Shuff-Bailey, Olguin, and Ruan (1995) performed a similar task whose results can be summarized by the paper's title: "Young children extend novel words at the basic level." Recent Bayesian approaches to vocabulary acquisition have replicated these effects of word extension from a single model, finding that children interpret the words as referring to basic-level or even subordinate categories (e.g., Xu & Tenenbaum, 2007). If global categories are primary, it is surprising that they are seldom the interpretation of the novel word in such tasks.

In their spontaneous word use, children largely avoid superordinate terms. Rosch et al. (1976) analyzed Sarah's utterances from Roger Brown's corpus during her earliest stages of language acquisition. She used 67 different basic-level terms, but only one superordinate term (*clothes*). Basic names were uttered 273 times, as opposed to two utterances of a superordinate. In single picture-naming, Rosch et al. (1976, p. 426) report that children used 269 basic labels and one superordinate label.[1] Anglin (1977) reports similar results.

When children are told that a familiar object has a new property, they tend to assume that the property holds for its basic category and not extend it to members of its superordinates (Waxman, Lynch, Casey, & Baer, 1997). Klibanoff and Waxman (2000) taught children new adjectives. Three-year-olds were able to generalize this adjective to objects in the same basic category as the learning item much more accurately than when the test object was only in the same superordinate category. The latter two results are particularly important in that they show that the basic category implicitly influences performance even when the task is not naming or categorization.

[1] Rosch et al. (1976) initially guessed that genus-level categories like oaks and robins would be the basic level, but discovered that higher-level categories like tree and bird were treated as basic instead. Nonetheless, they continued to test the same biological categories in their developmental work, sowing some confusion, since they did not actually test superordinate categories. I omit Rosch et al.'s biological categories from my summaries of their studies, because they cannot provide a fair test of the global-first hypothesis, since their highest level is the basic level.

A few studies have taught novel categories to children that were similar to basic- or superordinate-level categories. Horton and Markman (1980) created categories of novel animals and taught them to preschoolers and early elementary school children. They found better accuracy in learning the basic categories. Markman, Horton, and McLanahan (1980) also found that children through grade 8 had difficulty learning their superordinates (which were animate entities or abstract shapes) by ostension but had little difficulty learning basic-level categories.

2.1 Parental Input

Brown (1958) was perhaps the first to note that labels for an object differ in their abstractness and that parents tend to choose a middle level when speaking to children. Studies of parental speech have all found an enormous dispreference for using superordinate names. Blewitt (1983) found that in free speech to children, teachers and parents used only from 1−5% superordinate labels. Children heard far more subordinate than superordinate labels. In a story-telling paradigm (Study 4), college students used so few superordinates that they could not be analyzed. Speakers used more basic names when their story was directed towards a child. In picture-naming, Anglin (1977) found that mothers used basic-level terms the vast majority of the time, switching if necessary from the terms they would use for adults (e.g., saying *dog* instead of *collie*). They never used the familiar superordinate terms *jewelry, furniture, reptile, mammal, appliance,* or *vehicle* that would have fit the pictures. Nor did children use those terms.

Callanan (1985) found that when teaching superordinates, parents tended to also label the objects at the basic level. However, they did not label objects at the superordinate level when teaching basic-level names. Thus, parents expect that basic names would be helpful in learning more abstract categories, which would be more difficult for children to grasp.

The role of parental input can be seen as two-edged. Perhaps the apparent basic-level advantage children have is a simple result of their hearing basic-level terms more often. In its simple form, this hypothesis is not adequate, because some superordinate words are used more often than basic terms, simply because they refer to a larger class of objects (see Rosch et al., 1976, p. 426). A child is almost guaranteed to hear *animal* more than *zebra* or *ant*, and *food* more than *cake* or *hot dog*. Furthermore, correct categorization is not dependent on knowing the words. Rosch et al. (1976, Figures 2 and 3) showed that children's grouping performance was much better than their labeling of the categories they formed. For example, eight kindergartners and first graders sorted superordinates correctly in Experiment 9, but none of them accurately labeled the groups they made.

Finally, treating parental input as an explanation has a regression problem. Why do parents label the items at the basic level rather than at the superordinate? If frequency is the underlying cause, we must answer that it is because these parents heard their parents label things at the basic level. Why did their parents do so? Because *their* parents did so, etc. Eventually, one must produce an explanation for why this frequency difference occurred and has continued across generations. The obvious reason is that basic categories are easier for parents themselves to process, and they believe that these names will also be easier for children to understand. The first reason is clearly shown in the adult literature, and the experimental evidence on children's learning novel categories and interpreting words suggests that parents are right in thinking that basic-level names will be easier for children to understand. This is not to say that frequency has no effect, but merely that it cannot explain away the basic-level advantage shown over and over in developmental research, in tasks involving words and nonverbal responses, and with familiar and artificial categories.

2.2 Language Differences

Names for superordinates are often different from names for the more concrete levels. Rosch et al. (1976) documented that American Sign Language (ASL) had more conventional signs for basic-level categories than they did for superordinates. Later research by Newport and Bellugi (1979) reported that some superordinate labels in ASL are composed of conjunctions of basic-level names, abbreviated and strung together. For example, *vehicle* would be indicated by signs glossed as *car_plane_train_*etc. The exact basic-level names used in this conjunction are not necessarily fixed; they can vary across signers and occasions. Clearly, the names of basic categories must have been more essential and therefore come first in the development of ASL. Superordinate names could then be constructed on the fly from the already existing basic names.

Ellen Markman (1985) pointed out that superordinates in English are often mass nouns. For example, we can say *Bring a chair* but not *Bring a furniture*. Instead, we must use a classifier, such as *piece of furniture*. Superordinate labels *furniture, jewelry, clothing,* and *food* are syntactically like mass words such as *mud* or *rice*, which generally refer to undifferentiated conglomerations of similar stuff. In a study of 18 different languages, Markman found that superordinates across many different languages are much more likely to be mass nouns than basic names are.

It is more difficult to use a mass noun to refer to an individual object than to label it with a count noun (*Hand me the article of clothing* vs *Hand me the shirt*). Why would so many languages saddle their users with this less felicitous form of naming if superordinates categories are the most important and easiest to learn? Markman's explanation was that mass nouns make the superordinates easier to learn. If correct, this directly contradicts the assumptions of the global-first account that superordinates are the easiest to learn.

Researchers in ethnobiology have carried out a series of studies of how people in different cultures name the biological entities in their environment. Although it is impossible to briefly summarize the work of an entire field (see Malt, 1995; for an invaluable review), there are two findings that are particularly relevant here. One is that the neutral way to name objects in most cultures is roughly at the genus level, e.g., trout, elm, blue jay, rose, etc. (Historically, speakers of English did the same, before modern speakers lost most of their knowledge of specific biological categories and shifted up to terms like *fish* and *tree*; Atran, Medin, & Ross, 2005.) These preferred categories are two or more levels below those proposed as global categories in the infant literature (e.g., mammal, animal, plant). Furthermore, many cultures lack a simple name for those global categories—instead, a description or paraphrase is used to indicate animals or mammals as a whole (Berlin, 1992; Chapter 4). Berlin (1992, p. 190) summarizes research in ethnobiology as revealing that "in most systems" there are no names for the categories plant or animal. Basic-level names are said to be the first to enter the language (Berlin, 1972). So, in cultures without the influence of science and technology, the most useful categories are not the global ones, which may not even be lexicalized.

A number of metaphors are used to explain why the genus is so often named. One often-repeated one is that genera "cry out to be named." Berlin (1992, p. 10) says that "one segment of biological reality literally [!] jumps out at the viewer, something like a series of snow-covered mountain peaks on an exaggerated relief map. These peaks represent such obvious perceptual units as to be recognized almost automatically." If superordinate categories are the most important ones in our thought and easiest to acquire, why doesn't every language have a name for them and why aren't *they* the mountain peaks that are recognized automatically?

2.3 Conclusion

I have stuck primarily to older, "classic" studies in this section, in part to remind readers who may never have read them of the strength of the evidence for the advantage of the basic level in supra-infant conceptual development. I could have added considerably to this review by discussing the adult literature as well. There, most of the same phenomena are found but with response time usually showing the effect rather than errors. For example, adults are slower to identify something into a superordinate category than into a basic category (Maxfield & Zelinsky, 2012; Murphy & Brownell, 1985; Rogers & Patterson, 2007; Rosch et al., 1976), even though they are fairly accurate at both.

It is also worth noting that in Rosch et al.'s original paper, with 12 experiments, almost every study documented a basic-level superiority over superordinates, but a number did not demonstrate its superiority over subordinates. The basic-level advantage is more an advantage over superordinates than over subordinates. Thus, it is doubly surprising that in the infant literature, global categories that approximate those superordinates are learned as early as or earlier than basic categories. I turn to those findings next.

3. INFANT CATEGORIES

As explained in the introduction, the infant categorization literature has often found that infants are able to learn global categories, in spite of the great difficulties such categories seem to pose for the same children when they are 2 years older. This section will review those findings, focusing primarily on comparisons of category levels or demonstrations of superordinate categories in particular. Space does not permit a review of the entire literature of infant categorization.

3.1 Methods

The methods of studying cognition in infants have been well described elsewhere, and I will only give an outline of the relevant techniques used to study categorization. The dominant paradigm is the *familiarization-preference procedure*. Here, infants view a sequence of displays, usually drawings or pictures of items from a given category. At test, they are presented with two novel pictures, one depicting a new member of the familiar category and the other a member of a new category. For example, after viewing 12 pictures of different moose, the child might see pictures of a moose and a deer, presented simultaneously. The child's looking at each picture is measured. Preference for the novel picture (here, the deer) is taken as evidence for category formation. This effect requires two important components. First, the infant must have noticed some of the similarities of the initial pictures (the moose). Second, the infant must notice the difference between those pictures and the novel picture (the deer) and prefer to look at a novel stimulus. As a result, demonstrations of category learning in this paradigm can only speak to learning a category well enough to distinguish it from the novel category tested. For example, if we tested our infants on a moose versus a steam shovel, it would be going too far to describe this as the child having learned the category of moose. The child noticed something about the moose pictures that is different from steam shovels, but we don't know what. As the novel picture's category gets closer and closer to that of the learned category, we can say with more confidence that the child learned that specific category. If the baby could in fact distinguish moose from deer, that would suggest it learned the moose category quite specifically. Thus, I often report the specific discrimination tested (e.g., not that infants learned the category of dogs but distinguished dogs from birds).

The second popular technique, *sequential touching*, is used for older children who have manual control. Here, objects are presented on a tray or table, and children are instructed to "fix them up" or invited to play with them. The objects are generally models or toys representing members of two different categories, for example, half animals and half furniture. Children's touching of the objects is coded. If they show a tendency to sequentially touch more objects in a single category than would be expected by chance, this provides evidence that they can distinguish the two categories.

The logic behind the preferential looking task is clear, based on the well-attested novelty preference. Furthermore, the task is used successfully throughout the study of infant cognition, not just conceptual development. The logic of the sequential touching task is not as clear. Children are not instructed to separate the two categories; they can do whatever they wish with the objects. Why should they touch category members in order? Other goals they may have, such as playing with the objects or making them do different things, might well interfere with that measure. For example, if there are animals and vehicles, why shouldn't the child take the mouse for a ride in the boat rather than touch the mouse and then the dog and horse?

In spite of this skepticism, it must be said that there are multiple demonstrations of the sequential touching task revealing children's category discriminations. Apparently, it works. But as with many such techniques, positive findings showing that children do discriminate categories seem stronger evidence than null results showing they don't, as children's other goals or interests may have preempted their categorical touching. Informally, the results of sequential touching do not seem as consistent as those of preferential looking.

3.2 Infant Categorization Results

The first studies of infants' ability to form real categories were generally at the basic level. (I'm going to ignore important early studies using artificial categories, to avoid debating what level such categories are analogous to.) Cohen and Caputo (1978) and Roberts (1988), for example, showed that 12- and 9-month-old infants could learn categories of dogs and birds, respectively. Quinn, Eimas, and Rosenkrantz (1993) showed that 3-month-olds could distinguish cats or dogs from birds, using color photographs. Subsequent studies investigated when infants would learn distinctions such as cats from dogs, cats from horses, cars from airplanes, horses from fish, chairs from tables, and so on. Such categories differ on the perceptual features that are essential to the infant familiarization paradigm. Apparently, knowing the categories prior to the experiment is not necessary, as infants formed concepts of categories they no doubt had little contact with, such as airplanes and horses.

However, the assumption that infants' categories were confined to basic-level categories (or some approximation of them) was challenged by Behl-Chadha's (1996) dissertation. She first showed that 3-month-olds could make basic-level discriminations within artifact categories, such as chairs versus beds. She then examined global category learning and discrimination, testing children on mammals versus other animals (birds and fish) or artifacts (furniture). Her results showed that 3-month-olds could make such discriminations. A further study showed that children could distinguish furniture from mammals. There was insufficient evidence to show that the babies could distinguish furniture from vehicles.

Behl-Chadha (1996, p. 137) concluded that "a perceptually based process of superordinate-like categorization may begin and be operative in infancy at the same time that basic-level differentiations occur."

Quinn and Johnson (2000) attempted to raise the bar (or lower it?) by testing 2-month-old infants on global (mammal vs furniture) and local (cats vs other mammals) categories. They found evidence of global but not basic discrimination, which they concluded was consistent with a global- to basic- learning sequence. However, as evidence that global classification actually precedes basic classification, the results seem a bit weak. The effect for the global discrimination was not very strong ($t = 1.91$, $p < 0.05$, one-tailed), and the claim of precedence relies on null results for the basic discriminations. Since many experiments have shown basic category formation in 3-month-olds (e.g., see results of Behl-Chadha, 1996; above), it would seem to require more testing before one would be confident in saying that 2-month-olds cannot learn basic categories. However, as I will discuss later, the finding that global categories can be formed so young is very significant whether or not basic categories are also formed at that age.

Younger and Fearing (1999, 2000) used a slightly different logic to test basic and global classification simultaneously. Following the lead of Quinn (1987), they presented mixed pictures of objects from two basic-level categories to infants in the familiarization phase, rather than the more usual paradigm in which a single category is presented during learning. For example, one study used cats and horses. Then they tested whether the children formed the constituent categories as well as the overall category (mammal). This was done by contrasting a novel member of one of the constituents with a similar but different basic-level category (e.g., horse vs dog). They also tested a novel category that contrasted at the global level (for this example, car). They discovered that it was only at 10 months that the babies made both discriminations. At 4 and 7 months, they could make the global discrimination (mammal vs vehicle) but had not formed the separate categories (cat and horse).

This is an interesting and revealing result, but it is not the best design for examining whether global or basic categories are learned first, because the circumstances of their presentation are different. That is, the basic categories cats and horses were presented intermixed in the learning sequence, whereas the mammals were all presented together in the learning sequence. As Younger and Fearing (2000) point out, the presentation of pictures likely draws attention to common features, providing contextual support for the category that is represented in the entire set. That is, the task was harder for the basic categories than for the global categories, which were never presented mixed with a different global category. Nonetheless, their experiments provide another example of the formation of global categories as early as 4 months.

Mareschal and Tan (2007) were able to address the technical problems just described by using older children and a sequential touching technique. They presented multiple hierarchically organized categories at a time. For example, their display might contain boats and cars, both vehicles, plus cows and fish, both animals. Thus, children could simultaneously display knowledge of the basic-level and superordinate-level categories, which were both mixed in the display. In fact, about one-third of their 18-month-old children classified the items at both the global and basic levels. Overall, though, greater numbers of children showed evidence of basic categories, as 28% classified only at the basic level but only 8% only at the global level.

One important difference between the sequential-touching procedure and the familiarization paradigm is that it usually relies on currently existing concepts rather than testing the formation of a new concept. The exposure to toys on a tray is not enough to generate habituation (and the touching measure doesn't measure novelty preference). In contrast, presenting pictures of 12 elephants can teach infants something about the class of elephants even though they likely knew nothing about elephants prior to the experiment.

Jean Mandler and her colleagues performed a number of studies using the sequential-touching paradigm. Mandler had concerns that looking-based measures may reflect the formation of perceptual schemata, and not true concepts. She proposed that the sequential touching task engages conceptual knowledge (though it is not clear exactly why it should not reflect perceptual knowledge). Her articles have often reported that children are able to distinguish global categories prior to basic categories. Mandler, Bauer, and McDonough (1991) found that 19- to 30-month-old children could make a number of global discriminations: animals versus vehicles; land, air, and water animals; animals versus plants; kitchen items versus furniture. However, basic-level contrasts were only distinguished when the categories were fairly distinct (cars vs airplanes). Children could not distinguish cars from trucks, cacti from trees, tables from chairs, or spoons from forks.

Mandler and McDonough (1993) used a dishabituation paradigm with toy objects that children could examine (similar to the preference procedure, but with children tested on one object at a time). They found that 9- to 11-month-olds could distinguish animals from vehicles and cars from airplanes or motorcycles but not dogs from fish or rabbits.

Using a sequential touching paradigm, Rakison and Butterworth (1998) found that 14-month-olds could distinguish vehicles from animals, insects, and furniture. After modifying the stimuli by adding or subtracting parts (legs and wheels), the authors concluded that the basis for superordinate classification was largely perceptual, based on critical parts that infants tend to attend to. They did not test basic-level contrasts.

Overall, the results from the sequential touching research do not give a clear answer to which level of categorization is acquired first. Even within Mandler's results, some global contrasts weren't found (e.g., dogs vs fish, which should fall into land and water animals, which she identifies as global categories) and some fairly close basic contrasts were found (cars vs airplanes and cars vs motorcycles). Furthermore, some of the null results are very difficult to take at face value. Do 2-year-olds really not distinguish forks from spoons? Or tables from chairs? Given that these are the objects that children come into contact with every day of their lives, it's hard to believe that they don't know the difference between them. However, what can more positively be said is that children seem to be able to distinguish some global categories in this paradigm, probably as early as 11 months of age. That is surprising, given the literature on older children's problems with superordinates.

3.3 Animacy

A final issue that is sometimes mentioned in regard to the nature of early concepts is children's detection of animacy (e.g., Rogers & McClelland, 2004). Infants can distinguish natural biomechanical movements from unnatural movements as young as 3 months old (Bertenthal, Proffitt, & Cutting, 1984). Furthermore, animacy is apparently a feature of great interest to babies, who prefer to look at moving objects or images over static ones. However, this attentional preference does not necessarily reflect a concept of animacy. In a detailed analysis, Rakison and Poulin-Dubois (2001) point out that animacy is a complex construct embodying different perceptual cues and underlying assumptions. They conclude from their review (p. 221) that "In the 1st year, these properties are very much isolated from each other; for example, an infant might perceive causality and agency but not connect the two. We contend that there is little in the way of evidence that infants have developed a conceptual understanding of (i.e., know the meaning of) animates and inanimates before the middle of the 2nd year."

Babies are no doubt very interested in and learn much about animacy during their first years of life. However, there is no strong evidence that animate object is a concept that infants use to classify objects. In adults, at least, animacy is a *property*, not a category that objects are sorted into. It could well be that infants are attracted by cues to animacy without classifying objects as animate or inanimate. As discussed earlier, single-criterion categories like animate are not the same as rich, family resemblance categories that are used to classify everyday objects. Animate entities do not share shapes, parts, functions, colors, behaviors, sizes, etc. So, even though animacy is likely an important property to infants, it is unlikely to be a rich category.

3.4 Summary of Findings

This has by no means been a complete review of the field, but a more exhaustive review is not necessary, because the most critical conclusions are already clear. First, both basic and superordinate classifications have been found in both paradigms for the youngest infants tested. Whether 2-month-old infants can truly make superordinate but not basic discriminations is not yet certain. But more important is the fact that children have been credited with forming superordinates in multiple experiments, in both paradigms. All of those findings seem quite surprising from the perspective of the Roschian basic-level advantage.

As a result of these findings, multiple researchers have argued that conceptual development forms in a top-down order, in which large, global categories become differentiated into basic categories (fairly quickly), which may become further differentiated into subordinate categories, later on. For example, Quinn and Johnson (2000) refer to "Global-before-basic object categorization" in their title. Younger (2010, p. 254) summarizes her helpful review of this topic with, "we have seen that the pattern of responding to broad contrasts [i.e., global] at a younger age than narrower ones [basic] is evident in different task contexts across different age ranges." Mandler et al. (1991, p. 263) summarize their evidence as "by 18 months children have developed conceptual categories of animals and vehicles without yet clearly differentiating basic-level categories within these domains."

3.5 Summary of the Problem

The problem, then, is that the top-down differentiation of categories proposed in the infancy literature does not comport well with the difficulty that 2-, 3-, and 4-year-old children have with superordinate categories. Preschoolers don't seem to have acquired the global categories they were credited with knowing when they were infants, and they have great difficulty in learning analogous novel categories. In some cases, the tested categories that preschoolers have great difficulty with are roughly the same as those that infants can learn, e.g., mammal, animal, and vehicle. This decline in superordinate performance happens so early that we cannot even blame our educational system! What happened?

To answer this question, it will be useful to briefly describe the theory of why basic-level categories are easier to use than the other levels (Rosch et al., 1976; see Murphy, 2002, for detailed discussion). First, basic-level categories are associated with more features than superordinates are, at least if one requires features to have some reasonable frequency in the category. For example, chairs often have four legs, have a seat, have a back, are used for sitting on, have arms, and so on. However, none of these features is common to furniture as a whole. Indeed, when Rosch et al. tried to get naive subjects to list the features of different categories, they could not agree on a single feature that was common across tables, lamps, and chairs. Subordinate categories like armchair have even more features but also are very similar to other subordinate categories (dining room chair, reclining chair, side chair), and so classifying something into a subordinate becomes a difficult discrimination problem (Murphy & Brownell, 1985).

The features of superordinates are often rather abstract (Tversky & Hemenway, 1984). For example, musical instruments make music but don't all have keys or strings or legs; weapons are used to harm people, but this property is not a simple perceptual part like a blade or trigger. They create harm in different ways, with different parts. Such functional properties are not obvious from an object's appearance—seeing them requires practice and sophistication. In some cases, the critical property can only be perceived after categorizing the object, e.g., realizing that an object in a museum is a musical instrument prompts one to figure out how it makes sound. The shared properties of basic categories often include parts, which means that they often have a common shape as well (Tversky & Hemenway, 1984).

In short, if one thinks of classification as a matching process between a stimulus and a category representation, basic categories have the advantage of providing many matching features of the stimulus, and those features are easier to see than abstract or functional superordinate properties. Thus, it takes less time for that matching process to reach criterion (Murphy & Smith, 1982). At the subordinate level, there are many matches, but competing categories also have many matches: A stimulus that matches an armchair will also match the properties of a desk chair to some degree. Therefore, more information must be obtained to decide which category is correct.

These properties also should benefit the learning of basic categories. There are many features that independently predict category membership, so the learner doesn't have to immediately acquire them all. Those features are often concrete and easy to perceive, presumably helping learning as well. (However, the basic-level advantage that seems to occur with children and adults does not always occur in computational models of learning, as discussed below.)

My summary of the theory behind the basic-level advantage makes the point that we are not simply discussing an issue for which either answer could be correct without having much implication for our theories of cognition. If being a close match to a stimulus doesn't improve categorization, then how would we explain typicality effects (Hampton, 1995) or the time to make "different" responses in a same-different task? The principles that closer matches are better or that it's difficult to distinguish similar stimuli seem central to many aspects of perception and performance and to successful theories such as signal-detection theory. If the basic-level advantage is rejected, much work will have to be done to try to reconcile this with what we know about the rest of cognition.

4. MODELS OF INFANT CONCEPT ACQUISITION

Researchers have constructed connectionist models that learn conceptual structure from scratch. Such models provide a further way of understanding how infants may acquire their first concepts. A number of these models have discovered that global concepts are learned first, and these concepts are increasingly differentiated to form more specific concepts. Notable examples include Rogers and McClelland (2004) and Quinn and Johnson (1997, 2000).

I discuss these models in depth in Supplementary Material; space limitations preclude a complete discussion here. Different models have different strengths and weaknesses. Among the issues I point out regarding some of these models are unrealistic learning mechanisms, categories that are not based on empirically derived structures, and models that never appear to learn some of the categories taught, in contradiction to the infant studies. In a few cases, the models' basic-level categories actually seem to be acquired first. The most successful simulations are probably those of Quinn and Johnson (2000), in that they model actual stimuli used in an infant category-learning study and use a plausible learning mechanism. They report that the model learns global categories first. In the next sections, I discuss some of the shortcomings of the standard infant categorization studies, and that discussion will therefore apply to Quinn and Johnson's model based on those studies. Readers interested in more details should consult the Supplementary Material.

5. PROPOSED RESOLUTION

5.1 Identifying the Preferred Category Level

What is the nature of the basic-level advantage in adults? The fact that basic categories have many different structural and performance advantages makes it paradoxically difficult to define which one of them is necessary to make a category "basic." For example, the easiest way to identify the basic level for familiar categories is to discover which category is used to name objects in a neutral context. If one calls an object *book*, this suggests that book is the basic category, and not reading material or novel, because people use basic category names overwhelmingly in such tasks (Rosch et al., 1976; Experiment 10). But naming is not the only way one can tell such things, and one must imagine that some categories that don't have conventional names could still be basic categories. One would then have to fall back on category structure measures (Corter & Gluck, 1992) or other performance measures.

This issue is important because it must be addressed when comparing infant categorization to that of older children and adults. It is generally easy to compare even preschoolers to adults, because the adults can be given the same tasks (minus the puppets) as the children: naming, sorting, matching to sample, property induction, category learning, etc. However, no such measures can easily be made for infants. How then can we decide what categories are preferred by infants?

In the adult domain, the preferred level of categorization is supposed to be the neutral way that we identify and think about objects in our world. When we enter a room, we see chairs, a desk, a lamp, carpet, books, and so on. We do not see furniture, floor covering, and reading material. Some evidence for this claim comes from the finding that people very consistently name objects at the basic level, even when context suggests that more general or specific categories would be useful (Lin, Murphy, & Shoben, 1997). Furthermore, our behavior towards objects gives evidence that we have identified them more specifically than at the superordinate level. Although desks, chairs, and lamps are all furniture, we don't treat them identically: We put our papers on the first, sit on the second, and turn a switch on the third for light.

Infant behavior is such that it is difficult to tell whether and how the baby classifies an object in this way. Does the child see an object as being a cat? Clearly, the infant can't label the object as *cat* in order to communicate this. Unfortunately, its motor control and knowledge about the animal are minimal, so that we cannot generally look for cat-specific behaviors similar to my sitting on a chair and turning on the lamp. No doubt the infant has sets of expectations for the categories it knows, however. That is, after a couple of months experience with the family's cats, it expects the cats to meow and jump on furniture, but it does not expect a lamp to do either. In what follows, I will consider plausible arguments about what an infant might have learned in its environment and compare the results to conclusions from infant categorization studies.

5.2 Concepts Formed in Category-Learning Experiments

The infant categorization literature argues that its results give evidence that very young infants can form concepts such as mammals and furniture. However, I think it is fair to say that no infant has ever had the slightest conception of something being a mammal, even by fairly generous standards. The original studies are careful to specify that the infant has formed a "perceptual category" of the global category exposed in the familiarization phase (see below). However, mammal is not a perceptual category—that is why it took so long to be recognized in Western science. It is impossible to know that something is a mammal without knowing about the properties that are taught in elementary school: four-chambered heart, giving birth to live young, hair, warm-blooded, and so on. But either before or after being in a categorization experiment, infants have no ideas what any of these properties are, much less which animals possess them.

Because they don't know what these criteria are, infants cannot possibly separate mammals from similar nonmammals. The infant cannot know that dolphins, whales, otters, elephants, humans, mice, and bats are all mammals, but tuna, turtles, alligators, beetles, penguins, and passerine birds are not. Indeed, in many cultures, bats are classified together with birds (Berlin, 1992; Chapter 4), and in ancient times, whales were considered to be fish. Similarly, it would be surprising if an infant saw lamps, chairs, beds, and bookcases as all being the same kind of thing (furniture) but excluded appliances, stereo sets, street lamps, and trash cans. Without the knowledge of what those things do and why some are considered furniture and some not, there is no basis for making the distinction.

In infant concept acquisition studies, the items are fairly typical examples of the global categories. Behl-Chadha (1996, Experiment 3) used pictures of cats, deer, dogs, elephants, giraffes, hippopotamuses, horses, rabbits, squirrels, tigers, and zebras for her mammal examples. Missing from these 11 types of mammals are aquatic mammals, primates (human and nonhuman), or flying mammals. Most but not all of the pictures show the animal in profile, with four legs. Most have visible tails and ears. If children learn a perceptual category corresponding to these pictures, what is it? It probably includes mammalian facial features, plus an overall rectangular shape with four legs underneath. The contrast categories, birds and fish, differ in such properties and therefore would appear different to the infant. But one would not expect such an infant to now group gorillas, bats, or porpoises with the mammals, nor would one expect the infant to exclude alligators or dinosaurs.

Behl-Chadha (1996, p. 196) herself noted that the global discrimination she observed in infants was likely based on perceptual features. As she pointed out, most of her mammal pictures "clearly depicted four legs, distinct head and body regions, eyes, ears, tails, and so forth." Her furniture pictures tended to have long, straight vertical elements and straight edges, were vertically symmetrical, and had very regular shapes. Such features contrast with the properties of the mammal photos that babies distinguished from the furniture. Some of those furniture properties are also in many vehicle pictures (she tested cars and motorcycles), perhaps explaining why infants did not distinguish furniture and vehicles.

The genius of the familiarization-preference technique is that it can be used to test a wide variety of properties that infants might be sensitive to, not merely category membership or properties that would be used in category membership. It is used in many different areas of infant cognition and perception. For example, Behl-Chadha and Eimas (1995) showed that infants are sensitive to the left-right location of objects in this paradigm. After seeing pairs of horse and zebra photographs in a consistent order, they looked longer when the left-right order was changed. However, it is assumed that it is the structure of the displays themselves that draws children's attention to these properties (Younger & Johnson, 2011). Such findings cannot be taken as suggesting that infants are going around the world identifying the Cheerio as being to the right of the bowl, the cat as being to the left of the sofa, and so on. The experiments show that children *can* notice these properties under the right circumstances, but they are not interpreted as showing that children do consistently encode them when there is not this repetition and careful control of other differences. The baby is more likely to be focusing on the cat and what it is doing or is about to do than to be thinking of its left-right relation to the sofa. Similarly, I will argue below that infants may be able to see some kind of commonality to a set of pictured land mammals, which tells us something important about their classification abilities, but this is unlikely to be the way that they classify them when there is no learning repetition and preference phase to trigger the noticing of this commonality.

5.3 Research with Toy Models

The other main technique used to study infant categories uses models (e.g., toy animals and vehicles), either with the sequential touching measure or with a dishabituation measure in which children are allowed to engage with one object at a time. The dishabituation measure has properties similar to those of the familiarization-preference paradigm. For example, one could provide models that all have a point at their top. Then when given a new item with a smooth top, the child might well hold and examine it longer. But this would not entail that the child normally identifies objects as having or lacking points on top outside of this learning context.

The sequential touching paradigm is importantly different from the other ones in that it generally relies on actual concepts that the child may have prior to the experiment. For example, Mandler et al. (1991) put toy animals and vehicles onto a tray and invited the child to play with them. There is no learning phase here. The child's touching pattern is interpreted as reflecting his or her existing understanding of animals and vehicles. Mandler has interpreted the pattern of results in her work as indicating that children are sensitive to global categories prior to basic-level categories. I suggested above that the results here are actually rather confusing, as some basic-level categories are found quite early, and others that children almost certainly know (like spoon vs fork or table vs chair) are not found even after 2 years.

However, there is another issue with this paradigm that the research of Barbara Younger and colleagues has brought to light. The use of model stimuli to test concepts requires the subject to interpret the model as representing the actual object. That is, these studies are not asking whether babies can tell the difference between toys with legs and toys with wheels or between rough toys and smooth toys, they are asking whether children have concepts like animals, motorcycles, or fish. Younger and Johnson (2011) point out that young children have difficulty in relating realistic models to their referents, as shown in much work by Judy DeLoache (e.g., DeLoache, Pierroutsakos, & Uttal, 2003). The critical ages tested by Mandler and other researchers in the sequential touching paradigm are exactly those ages where children have these problems. Younger's research shows that 14-month-old children cannot match a model to a video of the same kind of object (e.g., cannot match a toy horse to a video of a real horse). That is, they apparently did not perceive the model as the same kind of thing as the real object. At 18 months, matching models to a real video was "fragile." Younger and Johnson suggest that learning about models of animals is to some degree separate from learning about animals, and that children can learn what the models are expected to do during play without making the connection that their referents do (some version of) those actions.

This is very interesting work with important implications beyond the methodological ones that are our concern here. For our purposes, these findings suggest that children's handling of model objects may not reveal the concepts they have of the actual objects depicted. Of course, sequential touching data may provide valuable information about what children see as similar and different in a number of respects. But Younger's research suggests that we should not assume that if children do or don't separate toy Xs and Ys, that means that they do or don't have concepts of X and Y for real objects. In the next section, I will also argue that the ways that models—and photographs—differ from real objects are consequential for concept formation.

5.4 Real-World Categories and Concepts

Let us imagine an actual infant, Renée, living on a peculiar kind of ranch where some of the diverse mammals used in infant categorization studies actually live. Let's suppose that on this ranch, Renée sees mice, cats, horses, elephants, and whatever birds happen to land in the yard. My question is whether the infant in this most interesting environment would first acquire and use global concepts like animal or mammal, instead of the relevant basic-level concepts.

In the infant categorization studies, infants see static pictures. Furthermore, the overall size of those pictures is controlled, so that the results are not due to a less interesting variable such as big versus small pictures (Behl-Chadha, 1996; Quinn et al., 1993, p. 466). Our infant, however, is not the beneficiary of this experimental control. The sizes of the mammals she sees differ by a factor of a 1000 or more. The mouse can barely be seen, the elephant can't possibly be missed, and the cat is clearly in between.

For college students, size is the most important factor in the sorting of different mammals (López, Atran, Coley, Medin, & Smith, 1997). If Renée is using mammal as the category to conceive of these entities, she is saying that the mouse and elephant are basically the same kind of thing, even though one could crush her, and she could crush the other.

The behaviors of the animals also differ greatly. The mouse darts in and out of holes and makes people scream. The cat sits for long periods of time waiting for the mouse but also runs and jumps inside the house. It sits and is petted. The horse stands and walks around outside; it never sits on its rear end. Occasionally someone rides it. The elephant lumbers around and is not ridden (at this ranch). The elephant eats by picking up hay with its trunk (a body part not found in the other animals). The cat and horse stick their heads into their food to eat; the mouse is not seen eating. The cat licks itself to be cleaned; the elephant is washed with a hose; the horse is brushed.

The sounds and smells associated with the animals differ greatly. The cat and mouse have no noticeable smell, but the horse and elephant have distinctive smells. The cat meows, the horse neighs, and the elephant trumpets loudly and makes Renée cry. The mouse is not heard. The colors and textures of the animals also differ. The horses are largely brown, the elephants and mice gray, and the cats are mixed white and orange. The cats have longish hair, the mice have short hair, the horse has a mane and short hair elsewhere, and the elephant is largely bald, with leathered skin. The faces of the animals are quite distinct, ranging from the prominent nose of the mouse to the huge ears and the trunk of the elephant to the relatively flat and whiskered face of the cat.

I have gone into more detail than one might expect to list the differences between these animals that we all know, simply to point out that their differences are very great, and the differences in size, movement and behaviors, and facial configurations would be highly salient to Renée. The elephant, horse, cat, and mouse are Berlin's "snow-covered mountain peaks" that jump out at you on the relief map. There is no way that Renée or any other human cannot notice the enormous differences among these basic-level categories. Furthermore, unlike the categorization study, Renée would be unlikely to see sequences of these animals one after the other. More likely she would be watching the cat for minutes at a time. When taken outside, she would no doubt stare at the elephant eating for minutes. When mommy is going for a ride on the horse, Renée might be in the barn, watching the horse being saddled, smelling and hearing it for an extended period. Thus, the commonalities that do exist in the mammal domain, such as presence of eyes and a nose, would not be reinforced by the constant switching of pictures. Instead, Renée might encode the particular shape of the cat's face and the haunches of the horse as it runs. Once heard, the sound of the elephant would not soon be forgotten. In contrast, a briefly presented picture of an elephant, followed by a picture of a mouse, followed by a dog, etc., all over the course of a few minutes, would provide a very different learning experience.

Upon first encountering each kind of animal, it has many very salient differences from the other animals, in size, body shape, and behaviors. Those differences only grow greater with more experience, as Renée learns what, where, and how the animals eat, for example, or as she sees them move or learns their different interactions with humans. It is true that some commonalities emerge as well: All of them see, all of them eat, all of them have legs, and so on. But some of those commonalities are abstractions that may not be readily encoded by an infant. Seeing the cat eat out of its bowl in the kitchen may not be perceived as the same thing as the elephant thrusting hay into its mouth with its trunk. Even if they are both understood as eating, the actions are very different.

My claim, then, is that the very large differences among these mammals would not be abstracted over by Renée to form a mammal category prior to forming more specific categories. When she sees one of them, she doesn't think "there's another mammal" or its infant equivalent. The elephant and the mouse do not appear to be the same kind of thing on almost any dimension in her experience. The horse and cat are also worlds different. When Renée sees the elephant, she doesn't expect to hear some generic mammal sound but expects to hear trumpeting; when she sees the mouse, she doesn't expect an average mammal movement but expects it to skitter away in mouse-like movements. If she hears the horse coming around the corner, she expects to see a large animal appear, not an average mammal-sized animal.

In contrast to real animals, the models of animals used in sequential touching studies are often quite similar in their size and perceptual qualities. The texture differences between animals or other objects are not well captured by plastic toys. They don't do anything; they make no sound; they don't differ in their smell. There is no worry about being crushed by the toy elephant; the toy cat won't give itself a bath. The same is true of photographs—they greatly reduce the actual differences between basic-level categories.

To be clear, this analysis is not intended to be a critique of the familiarization-preference paradigm, which provides critical data about infants' *abilities* in processing the presented stimuli. It tells us that babies can identify similarities in different pictures of cats or vehicles or horses. That is central to our thinking about what infants can do, and it is all the more remarkable given the poverty of the stimuli relative to real objects. But as a technique of telling us *what concepts infants actually have in everyday life*, it is less useful, because its stimuli and the nature of the experience are far removed from the learning experiences infants actually have. The positive actions of experimenters to improve experimental design by removing theoretically less interesting variables has paradoxically made the experiments less good at telling us what concepts infants form in the world where these things are not controlled. The extended nature of infants' experience with objects and the richer information in real objects must have an effect on what concept the child forms. (And lest readers feel I am picking on infancy researchers, I have made a very similar point about studies of adult category learning; Murphy, 2005).

5.5 Bridging from Infant to Child to Adult Concepts

So far I have questioned whether the finding of global categories in infant categorization studies gives strong evidence that infants form such categories from the objects they encounter in everyday life. I used a thought experiment of an infant who encounters a diverse set of objects from a single global category to argue that basic categories would be much more salient and likely to be formed. At this point, the argument is primarily a dispute in opinions: Researchers who perform infant categorization experiments could claim that their experiments provide reasonable evidence of what concepts infants actually form, whereas I argue that the properties that are absent from the typical infant category-learning study are exactly those that would promote basic-level categorization. In this section, I will argue that the basic-first proposal has a theoretical coherence that gives it an advantage over the global-first view.

Recall that there is very strong evidence for a basic-level advantage in adults and children from preschool through the school years. As I pointed out in my review, it is very puzzling that infants begin with global categories to interpret their world and yet cannot use those categories to group objects or make inferences a year or two later. If animal or vehicle or furniture are simple categories to learn because of their perceptual resemblance, why is that not a basis for categorization or induction when the child is a bit older? Furthermore, research on ethnobiology has argued that basic-level categories "cry out to be named," but higher-level categories do not, and hence are often not named in various languages. Why is it that infants don't hear those cries (according to the global-first view) but instead form categories that may not even be encoded in their native language, when they learn it? If superordinate categories have greater perceptual coherence (as the global-first theory must claim to explain the order of acquisition), why is it that adults are slower and less accurate at identifying objects at that level?

It is very difficult to construct a story in which infants start out at the global level, because it is in some way easier and more coherent, but then become unable over time to use those same categories when they must overtly classify objects or even implicitly, to learn a new name or generalize a property.

One response to this seemingly paradoxical developmental trend is to point out that concepts are not fixed entities. They change and grow with experience and especially over the years of childhood may undergo considerable evolution. This may be especially true of higher-level categories, which are often formed to some degree by scientific dictum. In our culture, children (eventually) learn that even motionless sponges or tiny gnats are animals in spite of their differences from the most common animals.

One could argue, then, that children start out with simple global categories that are in fact quite coherent. They are the result of grouping different entities that happen to share some observable features but may not share underlying structural properties that will eventually determine the adult superordinate. Quinn and Johnson's (1997) simulation picked up the fact that the mammals all had eyes and noses, whereas the furniture did not. Perhaps this is the kind of global category that the infant forms early on. When it develops further and also receives linguistic feedback, the infant discovers that these features are not reliable predictors of mammalhood. If the infant formed a category of wheeled vehicles, it then had to learn to incorporate boats and helicopters with very different shapes and no wheels. Thus, the vehicle category turned from a simple global category into a very difficult superordinate category that the young child may not have a strong grasp of.

This possibility is worth developing in future research. I think it is still difficult to explain why the initial global category based on perceptual features cannot be used to classify objects that share those features at age 2 and beyond (e.g., why learning a property applied to a rhinoceros, the child cannot identify it in a horse, Klibanoff & Waxman, 2000), but there is some plausibility to this story. However, we need to carefully consider exactly what those early "global" categories really are, under this view. If they are categories that are based on shared perceptual properties like eyes, ears, and a similar body shape (for four-legged mammals), then this is precisely a basic category. As explained in Section 3.5, basic categories are based on multiple shared properties, especially parts and overall shape (Rosch et al., 1976; Tversky & Hemenway, 1984). We have interpreted the infants' understanding of the experimental concept by reference to our own concept of mammals (or whatever), which was used in selecting the stimuli. But since such demonstrations generally use typical mammals that are perceptually similar to one another and altered to be the same size, they are not teaching "mammals," but the subcategory of mammals that share those properties. If this category is determined by multiple, shared perceptual features and overall shape, then it does not contradict the claim that basic-level categories are acquired first.

If there is a mistake to be found in the infant categorization studies, it is the idea of a perceptually based superordinate category. In order to make a superordinate category have significant perceptual commonality, one must edit it: remove size and movement differences, omit atypical members, and not test nonmembers that also have those properties. It is interesting that infants can form a category of this sort, but it would be wrong to then attempt to connect this category to a real superordinate, which has much more variability and difficult discriminations that require more advanced knowledge. The infant can't know that a street lamp, shown alone in a photograph, is not actually furniture, or that a swimming penguin and seal are essentially unrelated. Identifying photos of a camel and a horse as each displaying four legs in profile is not identifying a superordinate category. It may be the first step on a journey of learning the nature of superordinates, but the end of that journal will likely not be reached for years.

In sum, my conclusion is that the findings from the infant categorization literature have not yielded convincing evidence that infants are learning categories that are close to real superordinates. The superordinates have been simplified so as to become primarily perceptual categories, and the properties that are omitted from the experiments are those that would greatly favor the formation of basic categories. There is much to be learned from those infant studies, but applying their results to real-world categories is difficult.

My conclusion may, of course, be wrong. However, as I remarked at the beginning of this article, so far as I know, there is no developmental proposal of how infants start out with global categories and then turn into children who have so much trouble with superordinate categories. Nor is there a psychological account to explain why the features of basic-level categories that make them easier for adults would not apply to infants. If the global-first view is to continue to be held, its proponents need to grapple with the reliable findings of huge difficulties in processing superordinates in children and adults.

6. MORALS AND RECOMMENDATIONS

This review and analysis have raised a number of conclusions and issues deserving further study. First, if my conclusions are accepted, they confirm a view of conceptual development that it is a continuous process with many principles operating throughout the lifetime. I have argued that, contrary to Piagetian views, the data most strongly support the idea that infants are sensitive to the same learning variables as older children and adults: shared parts and shapes, distinctive categories, and sensitivity to the most coherent clusters of objects in the environment. The main difference between infants and older children is lack of knowledge and consequent inability to learn categories that depend on theoretical knowledge, such as being warm-blooded or being manufactured.

Another issue is that I believe that there is a strong relationship between cultural and linguistic markers of concepts and the psychology of concepts. I have argued that it is hard to understand why infants were thought to first acquire categories that in many cases do not exist in their (future) languages or cultures. Cross-cultural studies can reveal aspects of human cognition and life that are part of our human heritage, and infant cognition presumably reflects the same factors. It would be useful for cognitive psychology to consider in more detail the ethnographic literature when making proposals about what are essentially claims about universals of human cognition.

An additional issue is the tension that exists between laboratory studies and claims of concepts in the real world. This tension exists in the adult category-learning literature as well, to the degree that the study of category learning has very little contact with semantic memory research, which investigates the nature of our knowledge of the everyday world. In both cases, laboratory studies are best equipped to tell us about capacities and whether variables influence performance. However, to understand concepts in the real world requires an additional level of evidence, namely understanding the nature of real-world stimuli and learning conditions (Murphy, 2005). In the adult literature, there are many tasks to sample people's everyday concepts, such as category or property verification, memory tests, or association measures. The development of better methods for testing the *existing* concepts of preverbal infants would be of enormous benefit to the field.

Related to this is the issue that it is important to evaluate children's concepts on their own terms and not primarily in terms of the adult concepts that share their names. Researchers such as Carolyn Mervis and Jean Mandler have emphasized that children and adults have concepts that overlap but are not identical. In order to claim that children have a concept that is difficult for adults to learn or use, we need to test the concept carefully to make sure that children don't have a radically simplified version (e.g., meaning only four-legged mammals by *animal*).

A final point regards the difference between learning and other measures of category goodness. As a general rule, it seems that categories that are easy to learn also have performance advantages, such as fast categorization or resistance to forgetting. However, the infant concepts literature focuses on learning concepts, and the adult study of category levels focuses more on performance. This makes sense, because infants know so few concepts, and we cannot be sure which concepts an infant entering the lab might know. However, this difference has a specific implication for the understanding of category levels.

Computational models often seem to find that global categories are quickly learned, even though they have relatively few features that predict them. However, the benefit of the basic level in adult tasks seems to be largely the opposite result: More specific categories, with more features associated with them match stimuli better and so are faster. As people become experts, they become better at subordinate-level tasks, which are even more specific than basic-level categories (Tanaka & Taylor, 1991). When the issue of distinctiveness is removed, laypeople become fastest at classifying at the subordinate level (Murphy & Brownell, 1985). In both cases, superordinates remain the worst.

On the one hand, it makes sense that having fewer features (as superordinates do) is better—there are fewer things to learn. That seems to be what the computational models are picking up. But having many features leads to greater utility for a number of reasons. For example, if elephants have many features, one does not have to discover and focus on just one or two essential ones in order to start to acquire the category. The attentional requirements in learning may be therefore less for basic categories, an issue that is probably critical in early childhood. Furthermore, if some learned features are not perceptible in an object, others will still allow classification. I can still categorize something as an elephant if I don't see its trunk. If categorization requires one or two specific features to be observed, it cannot be very robust.

Specific categories also give you more information about the object. As I noted in the Supplementary Materials, if you knew that there was an animal in your back yard, you would know very little; if you knew it was a coyote, you would know a lot. However, the experiments and models we use to study categorization have focused on the classification aspect of concepts much more than this predictive aspect. Classification benefits much less from conceptual richness than prediction does. It is now widely accepted that concepts are shaped by the use to which one puts them (Markman & Ross, 2003). I think that progress in infant concepts would benefit from further thought about what uses infants have for their first concepts. Because of their lack of motor control and autonomy, the primary use is likely to be prediction. Therefore, discriminability may not be the most important determinant of what categories are first noticed and learned; it may be rather the richness or reliability of the predictions that one gains from the category. I offer this as speculation to stimulate future research into this question.

6.1 Concluding Thoughts

Research into infant cognition in the past few decades has arguably revealed more about human cognition than research in any other area of cognitive psychology, as new techniques provided discoveries of hitherto unsuspected mental processes. That investigation of course continues, including in the field of conceptual development. The seminal research into infant concepts performed by researchers such as Quinn, Eimas, Younger, and others has revealed much about infants' early abilities. Their ability to notice similarities and make discriminations is remarkable.

The limitations of the classic experimental paradigms must also be acknowledged. The familiarization procedure seems largely to reflect a learning process, in which the child identifies the commonalities of the pictures being displayed. It does not seem to depend much on the infant's experiences prior to coming in to the lab (Younger, 2010). Future clever researchers will have to find ways to discover babies' existing concepts to obtain more direct measures of at what levels they categorize objects.

Obviously, there is much more to be learned about infant and child concepts, and the present article can only hope to clarify one current problem, which will hopefully make future progress easier to achieve.

ACKNOWLEDGMENTS

Many thanks to Brian Ross and Karen Wynn for constructive comments. The writing of this chapter was supported in part by NSF grant BCS-1128769.

SUPPLEMENTARY MATERIAL

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/bs.plm.2015.09.002.

REFERENCES

Anglin, J. M. (1977). *Word, object and conceptual development*. New York: W. W. Norton.
Atran, S., Medin, D. L., & Ross, N. O. (2005). The cultural mind: environmental decision making and cultural modeling within and across populations. *Psychological Review, 112*, 744–776.
Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representation in early infancy. *Cognition, 60*, 105–141.
Behl-Chadha, G., & Eimas, P. D. (1995). Infant categorization of left-right spatial relations. *British Journal of Developmental Psychology, 13*, 69–79.
Berlin, B. (1972). Speculations on the growth of ethnobotanical nomenclature. *Language in Society, 1*, 51–86.
Berlin, B. (1992). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton, NJ: Princeton University Press.
Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology, 37*, 213–230.
Blewitt, P. (1983). Dog versus Collie: vocabulary in speech to young children. *Developmental Psychology, 19*, 602–609.
Blewitt, P. (1994). Understanding categorical hierarchies: the earliest levels of skill. *Child Development, 65*, 1279–1298.
Brown, R. (1958). How shall a thing be called? *Psychological Review, 65*, 14–21.
Callanan, M. A. (1985). How parents label objects for young children: the role of input in the acquisition of category hierarchies. *Child Development, 56*, 508–523.
Callanan, M. A. (1989). Development of object categories and inclusion relations: preschoolers' hypotheses about word meanings. *Developmental Psychology, 25*, 207–216.
Callanan, M. A., & Markman, E. M. (1982). Principles of organization in young children's natural language hierarchies. *Child Development, 53*, 1093–1101.
Cohen, L. B., & Caputo, N. (1978). Instructing infants to respond to perceptual categories. In *Paper presented at the Midwestern Psychological Association meeting, Chicago*.
Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: feature predictability and information. *Psychological Bulletin, 111*, 291–303.
DeLoache, J. S., Pierroutsakos, S. L., & Uttal, D. H. (2003). The origins of pictorial competence. *Current Directions in Psychological Science, 12*, 114–118.
Golinkoff, R. M., Shuff-Bailey, M., Olguin, R., & Ruan, W. (1995). Young children extend novel words at the basic level: evidence for the principle of categorical scope. *Developmental Psychology, 31*, 494–507.
Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language, 34*, 686–708.
Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51*, 708–719.
Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation*. London: Routledge and Kegan Paul.
Klibanoff, R. S., & Waxman, S. R. (2000). Basic level object categories support the acquisition of novel adjectives: evidence from preschool-aged children. *Child Development, 71*, 649–659.
Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin and Review, 3*, 95–99.

Lin, E. L., Murphy, G. L., & Shoben, E. J. (1997). The effect of prior processing episodes on basic-level superiority. *Quarterly Journal of Experimental Psychology, 50A*, 25—48.
López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32*, 251—295.
Malt, B. C. (1995). Category coherence in cross-cultural perspective. *Cognitive Psychology, 29*, 85—148.
Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford: Oxford University Press.
Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: differentiating global categories. *Cognitive Psychology, 23*, 263—298.
Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development, 8*, 291—318.
Mareschal, D., & Tan, S. H. (2007). Flexible and context-dependent categorization by eighteen-month-olds. *Child Development, 78*, 19—37.
Markman, E. M. (1985). Why superordinate category terms can be mass nouns. *Cognition, 19*, 31—53.
Markman, E. M., Horton, M. S., & McLanahan, A. G. (1980). Classes and collections: principles of organization in the learning of hierarchical relations. *Cognition, 8*, 227—241.
Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592—613.
Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: how level of target categorization affects visual search. *Visual Cognition, 20*, 1153—1163.
Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate, basic, and superordinate level categories. *Child Development, 53*, 258—266.
Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89—115.
Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
Murphy, G. L. (2005). The study of concepts inside and outside the lab: Medin vs. Medin. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the lab: Essays in honor of Douglas Medin* (pp. 179—195). Washington, DC: APA.
Murphy, G. L. (2010). What are categories and concepts? In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts* (pp. 11—28). Oxford: Oxford University Press.
Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 70—84.
Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior, 21*, 1—20.
Newport, E., & Bellugi, U. (1979). Linguistic expression of category levels. In E. S. Klima, & U. Bellugi (Eds.), *The signs of language*. Cambridge, MA: Harvard University Press.
Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition, 27*, 145—179.
Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*, 463—475.
Quinn, P. C., & Johnson, M. H. (1997). The emergence of perceptual category representations in young infants: a connectionist analysis. *Journal of Experimental Child Psychology, 66*, 236—263.
Quinn, P. C., & Johnson, M. H. (2000). Global-before-basic object categorization in connectionist networks and 2-month-old infants. *Infancy, 1*, 31—46.

Rakison, D. H., & Butterworth, G. E. (1998). Infants' use of object parts in early categorization. *Developmental Psychology, 34*, 49−62.
Rakison, D., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin, 127*, 209−228.
Roberts, K. (1988). Retrieval of a basic-level category in prelinguistic infants. *Developmental Psychology, 24*, 21−27.
Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
Rogers, T. T., & Patterson, K. (2007). Object categorization: reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General, 136*, 451−469.
Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382−439.
Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75* (13, Whole No. 517).
Tanaka, J. W., & Taylor, M. E. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 15*, 121−149.
Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*, 169−193.
Waxman, S. R., Lynch, E. B., Casey, K. L., & Baer, L. (1997). Setters and samoyeds: the emergence of subordinate level categories as a basis for inductive inference in preschool-age children. *Developmental Psychology, 33*, 1074−1090.
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245−272.
Younger, B. A. (2010). Categorization and concept formation in human infants. In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts* (pp. 245−263). Oxford: Oxford University Press.
Younger, B. A., & Fearing, D. D. (1999). Parsing items into separate categories: developmental change in infant categorization. *Child Development, 70*, 291−303.
Younger, B. A., & Fearing, D. D. (2000). A global-to-basic trend in early categorization: evidence from a dual-category habituation task. *Infancy, 1*, 47−58.
Younger, B. A., & Johnson, K. E. (2011). Emerging competence with symbolic artifacts: implications for the study of categorization and concept development. In L. Oakes, C. Cashon, M. Carasola, & D. Rakison (Eds.), *Infant perception and cognition: Recent advances, emerging theories, and future directions* (pp. 261−283). Oxford: Oxford University Press.

CHAPTER THREE

Believing that Humans Swallow Spiders in Their Sleep: False Beliefs as Side Effects of the Processes that Support Accurate Knowledge

Elizabeth J. Marsh[1], Allison D. Cantor and Nadia M. Brashier
Department of Psychology & Neuroscience, Duke University, Durham, NC, USA
[1]Corresponding author: E-mail: emarsh@psych.duke.edu

Contents

Psychology of Learning and Motivation, Volume 64
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2015.09.003

Abstract

Humans can store, maintain, and retrieve an impressive amount of information—but the processes that support accurate knowledge can also lead to errors, such as the false belief that humans swallow eight spiders in their sleep each year. In this chapter, we review characteristics of the knowledge base and explore how five adaptive properties that support accurate knowledge can also lead to the learning, storage, and retrieval of falsehoods. First, people exhibit a bias to believe information is true since, most of the time, incoming information is indeed true. Second, we utilize a fluency-based heuristic for judging truth since—again, most of the time—easy processing reliably signals that something is true. Third, the knowledge base is productive: people use existing knowledge to make new inferences, which are typically accurate but occasionally are inappropriate and result in errors. Fourth, existing knowledge supports new learning, so our ingrained misconceptions can foster new errors and interfere with learning the truth. Fifth, because it would be too taxing to carefully compare all incoming information to stored knowledge, we do not require a perfect match and often accept information as "good enough." As a result, errors that are similar to the truth often slip by undetected, and sometimes are later reproduced. Finally, we discuss methods for correcting errors and potential barriers to the correction of misconceptions. In particular, it is essential to refute the error as well as provide a simple alternative to replace it. Overall, the processes that support accurate knowledge and false beliefs are the same, and can lead to competing representations in memory.

1. INTRODUCTION

1.1 The Issue at Hand

The average person swallows eight spiders in her sleep every year.

Many of us have encountered some version of this claim. Is it true or false? Many people are unsure, but become concerned about the possibility;

a quick Internet search reveals many posts (e.g., on Quora, Reddit, Yahoo questions), where people explicitly ask others to verify this claim. Fortunately, science has provided absolutely no evidence to support this claim, and instead offers many reasons to doubt it (e.g., most humans move around a lot in their sleep; spiders avoid predators). So where did this idea come from? The original source is allegedly a journalist, who was poking fun at the ridiculous "facts" people learn on the Internet and unwittingly loosed the spider statistic on the world. In an ironic twist, this origin story may itself be an urban legend, as fact-checkers failed to locate the infamous article or even demonstrate that the author worked for the publication that supposedly published the piece.

The spider example may be laughable, but it demonstrates similarities between misconceptions and accurate knowledge. The cognitive processes used to encode, store, and retrieve veridical concepts (e.g., spider, sleep) form the basis for critical misunderstandings. We examine errors as "byproducts" of an otherwise efficient memory system and discuss ways to correct misconceptions after the fact.

1.2 Defining Knowledge

Stating the capital of Peru, solving a differential equation, and translating a text from Russian to English are all examples of successfully using knowledge. Knowing how to traverse airport security, deciding whether a joke is appropriate in a particular context, and calling one's sister by the correct name are also examples of successful use of knowledge. Defining "knowledge" is a tricky business, as the label applies to so many different things. Knowledge includes facts and concepts and an understanding of their relationships; knowledge also includes language, schemas, and inferences.

Most psychologists agree that knowledge is a form of memory extracted from past experience. However, knowledge is often defined by what it is *not*, rather than by what it is. That is, researchers contrast it to event memories (i.e., episodic memories; memories of specific events from particular places and times) with the emphasis on knowledge as information stored in memory that is decontextualized and that does not elicit a feeling of reliving. Depending on one's theoretical orientation, the term *semantic memory* may be considered synonymous with knowledge.

What is uncontroversial is the large role knowledge plays in many different cognitive processes. For example, knowledge drives how we interpret what we see. Imagine a photo of a girl standing on a balcony, looking at buildings. If we took a ruler and measured the image, the girl and the buildings might be the same height, but we do not interpret her as a giant or the buildings as miniatures; instead, we interpret their similar sizes as evidence that the building is further away than the girl, as we have learned from past experience. More generally, knowledge allows inferences, affects decision-making, guides the reconstruction of event memories, and supports communication and emotional responses.

2. GENERAL PROPERTIES OF THE KNOWLEDGE BASE

2.1 The Knowledge Base Has No Known Capacity Limit

There is no known limit to the amount of knowledge that can be stored in memory. Storage space does not become "full" over the years; to the contrary, older adults generally outperform younger adults on measures of vocabulary and general knowledge (Botwinick & Storandt, 1980; Mitchell, 1989; Perlmutter, 1978). As will be further described in Section 2.2, research on domain experts highlights just how much knowledge can be stored. Chess experts, for example, store an estimated 50,000 "game boards" in memory, allowing them to move quickly and automatically upon recognizing a particular board layout during a game (see Bedard & Chi, 1992; for a review). Such impressive memory feats are not specific to chess; expert musicians, bridge players, and computer programmers possess similar amounts of domain knowledge (see Ross, 2006; for a review). Computer simulations support these demonstrations of impressive knowledge; attempts to estimate the storage capacity of human memory by examining the rate at which people acquire new information suggest that we can store virtually limitless amounts of information (e.g., Landauer, 1986).

2.2 Knowledge Is Interconnected and Organized

Of course, knowledge does not consist of infinite separate pieces of information; that would imply that "the more one knows about a concept, the longer it would take to retrieve any particular fact about the concept" (i.e., *the paradox of the expert*, Anderson, 1983, p. 28). Instead, newly acquired information becomes integrated with existing information, creating an interconnected web of knowledge (e.g., Bartlett, 1932; Chi, Glaser, & Farr, 1988). This idea can be visually represented as a collection of "nodes," each of which represents a concept, with links to other related nodes (Collins & Loftus, 1975). The result is nonindependence among concepts; activating any one concept (e.g., by reading it, hearing it, etc.) "spreads" activation to other related concepts. Behaviorally, spreading activation manifests itself in *semantic priming*: people are faster to decide if a target (*nurse*) is a word after reading a related word (*doctor*) than after reading an unrelated one (*butter*) (e.g., Meyer & Schvaneveldt, 1971). This facilitation in reaction time, or *priming*, occurs because the concept of *doctor* was already partially activated after reading *nurse*. Activation continues to spread to concepts that are further away in semantic space, although the amount of activation decreases with semantic distance from the original concept. For example, *lion* can prime *stripes,* even though there is not a direct relationship between *lions* and *stripes*; *lion* primes *tiger*, and *tiger* primes *stripes*, meaning that exposure to *lion* yields observable priming of *stripes*, albeit less than for *tiger* (Balota & Lorch, 1986).

Of course, knowledge can be higher level than individual concepts, representing generalizations and extractions from past experience. A classic example comes from Thorndyke and Hayes-Roth (1979), where students read multiple texts describing different exemplars of a category. For example, a student might read about the constellations, Pisces, Aries, and Scorpio, before reading a target text about a new constellation. Memory for the target text depended on how many related passages preceded it; reading more passages boosted memory for the commonalities across passages, but this occurred at the expense of passage-specific details. In other words, participants extracted a *schema*, or generalized representation of constellation texts, which supported new learning at the expense of details. The more specific term *scripts* refers to action schemas, such as the steps involved in getting a haircut, shopping, and eating at restaurants. Supporting the existence of scripts, people are remarkably consistent when asked to generate the steps of common events like "eating at restaurant" or "getting a haircut" (Bower, Black, & Turner, 1979). Even though two strangers have never shared a restaurant meal together, they both know that a prototypical event begins with the hostess seating you, followed by the delivery of menus and ordering of food, and that at the end of the meal there is a bill and an expectation to tip. Both schemas and scripts are extracted from past experiences, and can be powerful tools for predicting outcomes in new experiences.

Expertise illustrates exactly how well-organized knowledge can be. That is, experts differ from novices in more than just the *amount* or *strength* of information stored in memory; expert knowledge differs qualitatively from that of novices in its *structure*. Chi and Koeske (1983), for example, compared a child's mappings of two sets of dinosaurs, one familiar to the child and the other unfamiliar. Unsurprisingly, the child's map for well-known dinosaurs boasted more structure than the one for unknown dinosaurs. That is, even though the novice map contained a similar number of property nodes (e.g., *eats plants*), the expert map contained many more linkages among dinosaurs, yielding a more interconnected and cohesive network (see also Gobbo & Chi, 1986). In addition, an expert's knowledge yields concepts that are more clearly differentiated from one another. For example, bird experts differentiate between warblers and finches more rapidly than do novices (Johnson & Mervis, 1997).

Experts also represent knowledge at a deeper level, whereas novices focus on surface similarities. Physics experts, for example, sort physics problems into groups based on principles of mechanics (e.g., problems pertaining to the Work–Energy Theorem) while novices group by literal features (e.g., problems with "blocks on an inclined plane"; Chi, Feltovich, & Glaser, 1981). A similar pattern emerges when people categorize fish; novices group by physical similarities across fish (e.g., groups for "all long and skinny fish"), whereas experts take functional information into consideration and form categories based on common habitats and for "fish that I eat" (Boster & Johnson, 1989). Furthermore, different types of expertise lead to distinct, but focused, organizations; for example, different tree experts sort trees differently depending on their particular expertise: maintaining city trees, landscape design, or science education/research. Both landscape designers and maintenance workers formed functional categories, whereas the scientists sorted the trees according to their actual scientific classifications (Medin, Lynch, Coley, & Atran, 1997).

2.3 Knowledge Is Surprisingly Durable

We forget many of our daily experiences, but knowledge proves surprisingly resilient, persisting across time and changes in context. In contrast to other kinds of memory, knowledge does not decline steadily with age; as mentioned earlier, older adults often outperform young adults on tests of general knowledge and vocabulary. Even after the onset of dementia, knowledge can sometimes remain intact and accessible. Hodges and Patterson (1995), for example, showed remarkable heterogeneity in the performance of patients diagnosed with minimal and mild Alzheimer's disease. All patients demonstrated event memory deficits, but some performed perfectly on measures of knowledge (e.g., category fluency: generating exemplars of a category).

Of course, much knowledge is encountered, accessed, and applied repeatedly over the years, effectively rehearsing the information. However, even knowledge that is *not* rehearsed appears to be remarkably durable over time. Knowledge of the Spanish language, the layout of a city one previously lived in, and names and faces of high school classmates all remain fairly stable over time, following an initial period of forgetting, even though participants do not report using or rehearsing the material since the time of original learning (see Bahrick, Hall, & Baker, 2013 for a review).

For example, Bahrick (1984) measured participants' retention of high school and college Spanish up to 50 years after initial learning. To estimate typical levels of acquisition, a subset of participants were currently enrolled in or had recently finished a Spanish course. The remaining participants had completed their last Spanish course between one and 50 years earlier. All participants took a number of recall and recognition tests for Spanish vocabulary, grammar, and reading comprehension. Performance on these tests was impressive: after an initial drop, the functions were quite consistent across the remaining interval. Based on these data, Bahrick argued that a portion of the knowledge base is so long-lasting that it is essentially permanent, or a *permastore* (see Figure 1 for a schematic based on hypothetical data).

Psychology instructors may be interested to know that a similar pattern occurs when examining students' retention of cognitive psychology course material (Conway, Cohen, & Stanhope, 1991). Since the majority of Bahrick's work examined the long-term retention of procedural knowledge (e.g., knowing *how* to speak a language), Conway and colleagues examined whether the same principles apply to declarative knowledge (e.g., knowing *that* cones allow color vision). To estimate students' retention of the

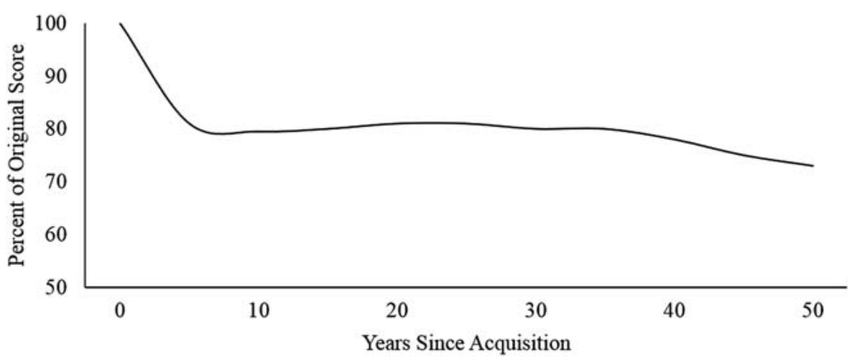

Figure 1 Hypothetical long-term retention of knowledge depicting Bahrick's (1984) "permastore."

material, the researchers administered a range of knowledge tests (tapping knowledge of both specific details and broader concepts) to students who had finished the course between 3 and 125 months prior. Consistent with Bahrick's findings, student performance on retention tests declined across the first 39 months of the interval, but then stabilized across the remaining portion.

2.4 Much, but Not All, Knowledge Is "Sourceless"

Retrieving knowledge and remembering an event "feel" different. While people "just know" that Washington, D.C. is the capital of the US, remembering a recent trip to the nation's capital evokes many associated details. Retrieving an event memory typically involves the feeling of traveling back in time and reliving the episode (Tulving, 1985). For example, remembering the events that occurred at a recent party might involve reexperiencing the music that was playing, the party decorations, the people in attendance, etc. In contrast, people often (but not always) fail to remember the source of their knowledge; when remembering that Washington was the first president of the United States, people do not normally think back to the original time and place of learning. In other words, people often report just "knowing" facts rather than "remembering" them. While knowledge may be linked to its source at first, this information is often lost over time, probably due to lack of rehearsal and to repeated encounters with the information that were associated with different sources. Supporting this claim, students initially judge course material as "remembered" but shift to "knowing" over the course of a single semester (Conway, Gardiner, Perfect, Anderson, & Cohen, 1997).

Consistent with these ideas, context appears to exert little or no influence on knowledge retrieval. For example, students do not always benefit from taking an exam in the same room where they attended class. While a few studies show small benefits of a contextual match (e.g., Abernathy, 1940; Metzger, Boschee, Haugen, & Schnobrich, 1979; Van Der Wege & Barry, 2008), others found no differences (e.g., Farnsworth, 1934; Saufley, Otaka, & Bavaresco, 1985). In contrast, there are many studies showing a benefit of contextual match for event memories, with people remembering more of a studied word list if they are tested in the same physical context. Overall, this benefit for event memories is "modest ($d = .28$) but reliable" (Smith & Vela, 2001, p. 213).

However, knowledge is occasionally associated with a source; these cases are often ones where there is a *reason* to remember the source—either because one will want to find the information again, or because the source might cast doubt on the veracity of the information. For example, there is some evidence that people are better at remembering *where* they found information on the Internet, as opposed to the information itself (Sparrow, Liu, & Wegner, 2011). Another example involves fictional sources, with the logic that readers/viewers should be hesitant to integrate everything from a fictional source into their knowledge base. In the study supporting this argument, subjects read a passage about the takahe bird that was labeled as *fiction* for some subjects but *factual* for others (Potts & Peterson, 1985). After reading the passage, participants made true–false decisions about the takahe either in blocks of other questions related to the takahe (passage context) or in blocks of unrelated questions (new context). Readers who believed the passage was fictional were slower to access their knowledge about the takahe when in a new context; no such effect occurred with readers of nonfiction. In other words, only information from the fictional story still retained some links to its original context.

2.5 Access to Specific Knowledge Can Shift

Storing and maintaining information in the knowledge base is not sufficient; just as important is the ability to retrieve that information when needed. Tulving and Pearlstone (1966) made the classic distinction between *available* and *accessible* memories: information stored in memory is available, but only information that can be retrieved is accessible. Although the availability–accessibility distinction comes out of the literature on event memory, the same idea applies to knowledge. People do not produce exactly the same knowledge at different points in time, reflecting the shifting accessibility of knowledge. Brown (1923) demonstrated this for the first time when participants attempted to recall the United States twice, 30 min apart in time. Even though participants' knowledge of the US states could not have changed over the course of the experiment, Brown found that participants forgot, or "lost," some states between the first and second tests and "gained" others. To get a better estimate of how much the accessibility of knowledge shifts across retrieval attempts, Bellezza (1984a, 1984b, 1984c, 1987, 1988) quantified the within-individual reliability of knowledge retrieval. His participants recalled as many category exemplars, noun meanings, facts about friends and family, scripts, and pieces of self-information as they could; critically, they did this twice, with the two attempts separated by 1 week. Across the various types of knowledge, reliability between two retrievals was modest but not very high; common-element correlations (McNemar, 1969) between two retrievals ranged from .38 to .69.

Bahrick coined the term *marginal knowledge* to describe knowledge that is stored in memory, but is currently inaccessible. Perhaps the most famous example is the nearly universal *tip-of-the-tongue* (TOT) experience, where one feels very confident that one knows a word or name or other piece of information but cannot produce it (e.g., Brown & McNeil, 1966; see Brown, 1991 for a review). TOT states likely reflect knowledge stored in memory that is available, albeit not accessible; people can frequently report the first letter or number of syllables of the target (e.g., Brown & McNeil, 1966; Yarmey, 1973) and such states often resolve with time (e.g., Choi & Smith, 2005; Read & Bruce, 1982). Presumably, TOT states resolve upon encountering different cues, something that could not happen if the targets were not actually stored in memory.

It can be difficult to distinguish the recovery of marginal knowledge from new learning. Berger, Hall, and Bahrick (1999) tackled this problem with a clever methodology, creating a set of fictitious questions that paralleled real ones. The fictitious questions matched the real questions in structure and sentence length, but had no factual basis (i.e., the researchers made them up). For example, for the real question *What is the name of the constellation that looks like a flying horse*? the parallel fictitious version asked, *What is the name of the constellation that looks like a sitting bull*? Critically, improvement on fictitious questions after an intervention (e.g., study phase) must reflect new learning, rather than reactivation of marginal knowledge. Berger and colleagues tested the ability of a 5-s exposure to stabilize access to answers that participants failed to produce on an earlier test. This intervention benefited real questions much more than fictitious ones, suggesting the existence of marginal knowledge for the real items. The benefits decreased with time; performance on an immediate test (90%) dropped continuously over 9 days (to 49%, see Table 1). In our own work, we showed that answering a multiple-choice question effectively reactivates marginal

Table 1 Proportion of final test questions answered with targets for real and fictitious questions (given initial test failure)

	.83 min	1.68 min	5 min	20 min	1 day	9 days
Real	.90 (.01)	.89 (.01)	.81 (.01)	.72 (.02)	.68 (.02)	.49 (.02)
Fictitious	.66 (.01)	.60 (.01)	.44 (.01)	.31 (.01)	.14 (.01)	.03 (.00)

Note. Standard errors are presented in parentheses.
Table adapted from Berger et al. (1999)Table 1.

knowledge (Cantor, Eslick, Marsh, Bjork, & Bjork, 2015), and such questions need not be paired with feedback.

2.6 People Are Good, but Not Perfect, at Judging What They Know

People generally have a good sense of whether or not they know something (see Nelson, 1988 for a review). In a classic example, Hart (1965) examined the accuracy of *feeling-of-knowing* (FOK) judgments. Participants answered a series of general knowledge questions; when they could not answer a question within 10 s, they judged whether or not they could successfully recognize the correct answer among several wrong answers. Participants' FOK judgments accurately predicted their stored knowledge: when participants claimed to hold knowledge, they correctly recognized the target 66% of the time.

Conversely, people demonstrate awareness of what they do *not* know. Hart (1965) found that when participants judged that they did not know the answer, they subsequently failed to select the target 62% of the time. Furthermore, Glucksberg and McCloskey (1981) proposed two different types of *don't know* judgments. The first is the slow, low-confidence *don't know* made in response to questions like *Is Kiev in the Ukraine?* The other is the fast, high-confidence decision made when asked questions like *What is Jimmy Carter's favorite color?* These two *don't know* judgments differ because of the amount of related knowledge stored in memory. When people have some knowledge about the topic (e.g., Ukraine), they search memory for the target answer. However, when people know little to no related information, no search occurs, resulting in a quick *don't know* response. In line with this dichotomy, Glucksberg and McCloskey found that people responded *don't know* more quickly to questions like *Does Bert Parks have a degree in journalism?* than to questions where the participants presumably drew on some relevant knowledge (e.g., *Does Anne Landers have a degree in journalism?*).

Of course, people's judgments of what they know are not perfect; the classic example is the *hindsight bias*, whereby people claim to have "known it all along" when told the answer to a question they could not answer (see Roese & Vohs, 2012 for a review). Similarly, people sometimes claim to know about nonexistent topics, for example, reporting use of fictitious products (i.e., *overclaiming;* e.g., Phillips & Clancy, 1972). Why do people overclaim? While multiple factors are likely involved, recent work highlights the role of self-perceived domain knowledge (Atir, Rosenzweig, & Dunning, 2015). In one experiment, participants rated their general knowledge about finance compared to the average American, then rated their knowledge about specific financial concepts. While some of these topics were real (e.g., tax bracket), the researchers fabricated others (e.g., pre-rated stocks). Finally, participants took a financial literacy quiz to estimate their actual domain knowledge. Higher self-perceived knowledge predicted overclaiming, independent of participants' actual domain knowledge. Furthermore, this effect appeared to be domain-specific, in that self-perceived expertise in a given domain related specifically to overclaiming in that domain and not others.

3. EXAMPLES OF ERRORS

3.1 Overview

Knowledge is impressive but not perfect; it is virtually unlimited in capacity and lasts for years, but it is not always available when one needs it. Gaps in knowledge are not surprising; what is more interesting than simple errors of omission are errors of commission. That is, people also believe things that are not actually true: errors can be stored in the knowledge base. For example, many people believe that we only use 10% of our brains, that the crime rate increases during the full moon phase, that seasons reflect differences in the physical distance between the Earth and the Sun, and that raindrops are tear-drop shaped. Misconceptions arise across domains; people hold false beliefs about science (e.g., McCloskey, 1983; Munson, 1994; Nakhleh, 1992), health (e.g., Lee, Friedman, Ross-Degnan, Hibberd, & Goldmann, 2003; Wynn, Foster, & Trussell, 2009), and probability and chance (e.g., Ferland, Ladouceur, & Vitaro, 2002), among many others. We will now describe a few examples that have both real-world parallels and laboratory analogs, before turning to some general principles that help explain *why* these errors occur.

3.2 The Grading Problem

Most educators have experienced the unfortunate feeling of becoming "dumber" after grading error-ridden exams and papers. Brown (1988) captured this phenomenon experimentally, examining how exposure to spelling errors hurts one's ability to spell correctly. After checking that participants knew how to spell the target words, they read or generated misspellings; later, they spelled the words again. People tended to switch from a correct to an incorrect spelling on the final test after seeing errors in the interim. This finding is striking, given that participants had likely spelled and read the correct version hundreds of times in the past. Why does a professor's brief exposure to a spelling mistake matter so much?

3.3 Side Effects of Reading Novels and Watching Movies

Movies and novels often are set in real places, refer to actual objects, and occur in familiar time periods, so they constitute a source of information about the world. For this reason, educators sometimes incorporate fiction into their course materials to better engage the students in learning (e.g., Dubeck, Bruce, Schmuckler, Moshier, & Boss, 1990). However, by definition, works of fiction contain inaccuracies, and as such have the potential to serve as sources of misinformation about the world. For example, viewers pick up errors from historically inaccurate portrayals in films and furthermore often misattribute that information to a historically accurate text (Butler, Zaromb, Lyle, & Roediger, 2009; Umanath, Butler, & Marsh, 2012). In one study, participants first read a historical text about the Satsuma Rebellion and then watched a clip from the popular film, *The Last Samurai*. While the text accurately stated that the Emperor Meiji hired a French military advisor to help quell the rebellion, the film inaccurately portrayed the advisor's nationality as American. Even after instructions to rely only on their memory of the text, participants answered questions (e.g., *From what country did Emperor Meiji hire military advisors*?) with inaccuracies depicted in the films (Butler et al., 2009).

We captured the fiction reader's experience in the lab by giving participants short stories with characters, dialogue, and plot. Critically, each story contained references to facts (see Marsh, 2004 for materials). The references were correct (e.g., *paddling around the largest ocean, the Pacific*), neutral (*paddling around the largest ocean*), or misleading (e.g., *paddling around the largest ocean, the Atlantic*). Participants later took a general knowledge test that probed the critical facts from the stories (e.g., *What is the largest ocean on Earth*?). In multiple experiments, reading misinformation (e.g., *Atlantic*) dramatically increased students' production of that error on the final general knowledge test. This effect occurred even after explicit warnings that authors of fiction often take liberties with the truth (Marsh & Fazio, 2006) and drawing attention to the errors with text signals (Eslick, Fazio, & Marsh, 2011). Why do students continue to rely on such low-credibility sources?

3.4 Repeated Claims Feel True

Used car salesmen, politicians, advertisers, and carnival barkers all capitalize on one truism: repeating something makes it seem truer. In the laboratory, falsehoods like *Zachary Taylor was the first President to die in office* appear truer if they were also seen earlier in the experiment. This phenomenon, coined *illusory truth* (Hasher, Goldstein, & Toppino, 1977), generalizes to political opinions (Arkes, Hackett, & Boehm, 1989) and claims about consumer products (Hawkins & Hoch, 1992). The influence of repetition persists minutes (e.g., Begg & Armour, 1991), weeks (e.g., Bacon, 1979), and even months (e.g., Brown & Nix, 1996) after initial exposure to the claim. Reminding participants that claims come from untrustworthy sources reduces the effect (e.g., Begg, Anas, & Farinacci, 1992); as mentioned earlier, though, people often fail to monitor source. Henkel and Mattson (2011), for example, found an illusory truth effect for statements that participants later identified (correctly or incorrectly) as coming from an unreliable source. Why do the repeated statements made by politicians or salesmen influence us, even when voters and consumers realize that they may not be credible sources?

3.5 Tests Can Teach Errors

Considerable controversy surrounds the educational value of tests; concerns include teaching to the test and decreasing motivation in students. Another possible side effect involves the potential of some tests, particularly multiple-choice tests, to introduce errors into the knowledge base. Multiple-choice tests by definition pair a correct answer with multiple plausible, but incorrect, answers. In other words, such tests expose students to more errors than correct answers.

Across multiple studies, people reproduced some of the multiple-choice lures from an initial test on later tests (e.g., Bishara & Lanzo, 2015; Fazio, Agarwal, Marsh, & Roediger, 2010; Odegard & Koen, 2007; Roediger & Marsh, 2005). In one study, students took a multiple-choice test consisting of retired SAT II questions about biology, chemistry, U.S. history, and world history (Marsh, Agarwal, & Roediger, 2009). The correct answer and four lures accompanied each question; participants also had the option to skip questions. After a short delay, participants took a final general knowledge test consisting of short-answer questions; some of these corresponded to the earlier multiple-choice items, whereas others were new. Prior testing helped overall: participants were more likely to answer correctly if a question had appeared on the earlier multiple-choice test. However, recent exposure to multiple-choice questions also increased the probability of incorrectly answering the questions with lures. This effect diminishes over a delay and with feedback (Butler & Roediger, 2008; Marsh, Fazio, &

Goswick, 2012), but increases if students rehearse the multiple-choice lure on a later test. Why do students pick up errors from assessment tools that otherwise boost learning?

4. ADAPTIVE PROCESSES THAT ALSO SUPPORT ERRORS

4.1 Overview

This section focuses on the mechanisms by which errors enter the knowledge base with implications for how to correct them (see Section 6). It is easy to identify potential sources of errors: we could point to textbooks that unfortunately contain errors (e.g., Cho, Kahle, & Nordland, 1985), the learner's own illogical reasoning (e.g., Clement, Narode, & Rosnick, 1981), and other people such as family and friends (e.g., Landau & Bavaria, 2003). Knowing that errors exist in the real world, however, does not tell us why people accept them as facts, reproduce them later, and let them influence their beliefs. For example, how do incorrect portrayals of amnesia in films (e.g., Uncle Fester in *The Addams Family*, Dory in *Finding Nemo*) contribute to people's misunderstanding of amnesia and traumatic brain injury? When viewers simply know nothing about neuroscience, it is unsurprising that they rely on such depictions. However, as discussed below, the problem extends beyond mere naiveté. Below we consider five interrelated properties of how knowledge is encoded, stored, and retrieved; all are properties that normally support accurate knowledge, but sometimes backfire and allow the introduction of errors into the knowledge base.

4.2 Property #1: Bias to Believe Information Is True

Daily life barrages people with new information, some true and some false. How do people decide the truthfulness of claims in the environment? Do they automatically know the truthfulness of statements like *The Pacific is the largest ocean on Earth* and *Lexington is the capital of Kentucky*? One argument is that comprehending a statement requires automatically accepting it as true; "unbelieving" involves a second, resource-demanding step. Gilbert reintroduced this idea to psychology in the 1990s, borrowing from the philosopher Baruch Spinoza. He illustrates the automatic acceptance of new information with a library analogy: the librarian assumes all books to be nonfiction *unless* they are marked by a special "fiction" tag. The reader may wonder why a librarian would ever take this approach to sorting books; put simply, it saves time and resources. It would take a librarian a lot more time to tag every book as fiction or nonfiction, as opposed to simply tagging the fictional ones. This strategy makes sense in the real world as well, where truths occur more frequently than falsehoods.

In most situations, automatically accepting claims conserves time and energy. However, strains on cognitive resources, like competing demands on one's attention, may prevent readers from actually reappraising and "unbelieving" false claims. Gilbert, Krull, and Malone (1990) demonstrated this phenomenon experimentally by blocking the evaluative, unbelieving step. Participants first learned fictional statements (e.g., *A twyrin is a doctor*) that were explicitly labeled as "true" or "false." While reading some items, participants performed a second task, which presumably interrupted the unbelieving step. In a second phase, participants judged the truth of the claims seen earlier. Compared to reading statements alone, distraction led participants to make more "true" judgments later. In other words, participants never reached the stage of evaluating and tagging false statements.

Critically, Gilbert, Tafarodi, and Malone (1993) replicated this effect with real-world judgments. Participants read two crime reports, each containing both true (black font) and false (red font) information about robberies, with instructions to evaluate the information carefully; they knew that they would later play the role of a judge. In one report, the false information exacerbated the crime (e.g., *The robber had a gun*); in the other, the false information extenuated the crime. Half of the participants proceeded through the reports uninterrupted while the other half completed a second task while reading. Later, they decided how long the prison term of each perpetrator should be (0–20 years). Participants who read the reports under full attention assigned similar sentences to the crimes exacerbated and extenuated by false information. Interrupted participants, however, incorporated false information into their judgments: the prison terms for exacerbated crimes were twice as long as those for extenuated crimes. Disturbingly, this pattern emerged even after a strong warning to pay attention to detail and with false information explicitly marked. Even when preparing to make consequential decisions, people initially accept claims as true. This "economical" and "adaptive" bias (Gilbert, 1991) ultimately leaves us vulnerable to errors.

4.3 Property #2: Fluency-Based Heuristic for Judging Truth

As explained in Section 2.4, people often experience knowledge as information they "just know" without necessarily "remembering" where they first learned it (e.g., Tulving, 1985). Instead of thinking back to a particular time and place, people judge their knowledge based on how easy or hard it is to retrieve information. Assuming one knows the capitals of France and Turkey, Paris likely comes to mind more easily than Ankara. This experience of *retrieval fluency* is in turn interpreted as confidence in one's answer.

Kelley and Lindsay (1993) experimentally demonstrated that retrieval fluency causes "illusions of knowledge." Participants read a list of words before completing a general knowledge test. Critically, some of the studied words were semantically related to the answers to the subsequent test questions. For example, participants studied *Hickock,* a reference to the American folk character "Wild Bill" Hickock, and then later answered a question about a different folk character with a similar name: *What is Buffalo Bill's last name? Hickock* came to participants' minds easily due to the recent exposure; people answered the general knowledge questions with related, but wrong, answers because they were easy to retrieve. People made these errors with high confidence, demonstrating that their responses were not just guesses. In other words, participants mistook retrieval fluency for actually knowing an answer.

In our own work, we showed that this illusion can occur even if people remember the source of the misinformation. That is, people read stories containing errors (e.g., *St. Petersburg is the capital of Russia*), which increased their likelihood of answering later general knowledge questions with misinformation (e.g., answering *What is the capital of Russia?* with *St. Petersburg*). Critical for present purposes is that readers made two judgments about each answer; first, they indicated whether or not they had read each answer in one of the experiment's stories and second, they indicated whether or not they had known the answer prior to coming to the experiment. The bottom line is that readers were good at remembering the story sources, but they *also* misattributed these answers to pre-experimental knowledge (Marsh, Meade, & Roediger, 2003). This finding likely reflects how information in the real world is often encountered in multiple contexts—meaning that remembering a lower credibility source does not preclude information from also being associated with a reliable pre-experimental source. This illusion of prior knowledge even occurs in young children (Goswick, Mullet, & Marsh, 2013).

More generally, the ease with which we process information (i.e., fluency) serves as an extraneous cue for many judgments, including truth; perceptions of truthfulness increase when information pops to mind or even when statements are easy to read. Because **Antananarivo is the**

capital of Madagascar is much easier to read than Antananarivo is the capital of Madagascar, the former seems more truthful (Reber & Schwartz, 1999). As mentioned earlier in Section 3.3, illusory truth also occurs with repetition; repeated statements are easier to process, and thus receive higher truth ratings (see Dechêne, Stahl, Hansen, & Wänke, 2010, for a meta-analysis).

This heuristic proves to be both cognitively inexpensive and effective, as fluency naturally correlates with truth (Unkelbach, 2007). On average, the single true version of a statement (e.g., *The capital of Argentina is Buenos Aires*) occurs more frequently in the environment than any one of its many possible falsifications (e.g., *The capital of Argentina is La Paz, The capital of Argentina is Lima, The capital of Argentina is Montevideo,* etc.). People learn this relationship between truth and fluency with experience, as relying on fluency typically leads to the correct judgment (Unkelbach, 2007). In the absence of timely and accurate feedback, as is often the case in real life, people accept fluent errors and incorporate them into their knowledge bases. This tendency renders people vulnerable to misinformation in repeated advertisements, political propaganda, and rumors.

As many advertisers seem to understand, repetition is not the only means of creating a fluent experience. Pairing a statement like *The first windmills were built in Persia* with a photograph of a windmill in an unidentifiable field increases truth ratings. This effect, coined *truthiness*, occurs despite the fact that the picture provides no further evidence for the specific claim (Newman, Garry, Bernstein, Kantner, & Lindsay, 2012). Several mechanisms likely contribute to the power of pictures; for example, the photograph of a windmill may encourage people to generate pseudo-evidence for the claim (e.g., "this field looks arid, so maybe it was taken in Persia"). Among these candidates, fluency is bolstered by the most empirical evidence. Critically, truthiness only occurs when people view a mixture of statements, some paired with pictures and others appearing alone (Newman et al., 2015); in other words, statement–picture pairs seem truer only when contrasted with statements appearing alone, which are presumably less fluent. This result parallels the finding that illusory truth only emerges when people rate a mixture of repeated and new statements.

4.4 Property #3: The Knowledge Base Is Productive

Not all information incorporated into the knowledge base needs to be directly encountered in the outside world. People go beyond the information stored in memory to generate new knowledge. Consider a simple example, whereby 6-year-old children successfully integrated facts to arrive at novel inferences (Bauer & San Souci, 2010). Children learned that *Dolphins live in groups called pods* and *Dolphins communicate by clicking and squeaking*, and then later demonstrated knowledge that *Pods communicate by clicking and squeaking* (which was never explicitly stated). Electroencephalography in adults suggests that newly inferred facts possess a phenomenology similar to that of facts learned long ago: well-known and integrated facts resulted in similar P600 responses, which reflect the ease with which information is processed (Bauer & Jackson, 2015). This positive activity peaking at 600 ms also reflects whether information is stored in long-term memory. In other words, people readily generate inferences that "feel like" facts they learned years ago.

This remarkable ability allows us to bridge gaps but also has the potential to introduce errors into the knowledge base. For example, consider the reader faced with the following passage:

> *That's why we had to go to St. Petersburg, but at least I got to see the Kremlin while there. Her family came too—even though they lived in Russia's capital city, they had never visited the Kremlin!*

What happens when the reader is later asked *What is the capital of Russia?* The passage incorrectly implies, but never explicitly states, that the capital is St. Petersburg. Butler, Dennis, and Marsh (2012) demonstrated that inferences formed while reading persist, leading participants to reproduce errors on a later general knowledge test. In addition to forming incorrect inferences following misinformation, people may self-generate errors by false analogy or other misapplications of logical processes.

People also generate inappropriate inferences when they can retrieve knowledge relevant to one, but not both, objects of a comparison. When presented with the question *Which city is larger, Milan or Modena?* most people respond *Milan* simply because they recognize its name. After successfully retrieving knowledge about Milan and failing to retrieve knowledge about Modena, people arrive at the inference that Milan is larger. This *recognition heuristic* extends beyond judgments of city size: people infer that a recognizable option scores higher than an unknown option on any criterion.

The recognition heuristic provides a cognitive shortcut, allowing people to make a judgment where they otherwise could not. As is the case for fluency, relying on recognition leads to accurate judgments in many cases, even when pitted against more complex approaches. When asked to judge the relative sizes of German cities, for example, American students outperformed their German counterparts, while the reverse pattern emerged for American cities (Katsikopoulos, 2010). Here, using additional cues actually harmed performance (i.e., *less-is-more effect*). However, recognition can be misleading in plenty of situations; for example, many people would incorrectly answer *Which city is larger, Hong Kong or Chengdu?* with *Hong Kong* on the basis of recognition alone. A remarkable proportion of people (nearly 50%) consistently base their judgments on recognition even when faced with three contradicting cues (e.g., learning that a recognized city does *not* have an international airport; Pachur, Bröder, & Marewski, 2008). In addition to leading people astray at the time, the errors generated during these comparisons may persist over time.

4.5 Property #4: Existing Knowledge Supports New Learning

Learners do not enter new situations as blank slates; they bring along a wealth of knowledge, from specific concepts to generalized structures (see Section 2.2). Most of the time, this knowledge supports understanding of, and later memory for, novel concepts. Experts generally learn new information within their expert domain more readily than novices, because they can scaffold and integrate this new learning with an impressive foundation. For example, Van Overschelde and Healy (2001) looked at the ability of baseball experts and nonexperts to learn fictitious statements about real baseball players (e.g., *Sammy Sosa likes to read books by John Steinbeck*); the fictitious materials ensured that participants could not have these details stored in memory. Even though the facts were not relevant to the game, baseball experts outperformed novices on a memory test.

However, knowledge can also interfere with new learning, particularly when used inflexibly. For example, children typically learn that a *particle* refers to a specific entity (e.g., a particle of sand), which then interferes with their ability to grasp a different conceptualization of particle, as used in physics. This problem is an example of *linguistic transfer*, whereby an established use of a word interferes with learning a new usage (Clancy, 2004). Another example involves the simple terms *while* and *then*, which take on different meanings for computer programming than they do in everyday English. Specifically, the everyday usage of the word *then* has temporal implications, which interfere with thinking about *then* as a conditional (e.g., Bonar & Soloway, 1985).

Similarly, knowledge interferes with problem solving when the rules change. For example, *functional fixedness* occurs when a learner cannot think of a new way to use a familiar object. In the classic candle problem, Duncker

(1945) instructed participants to attach a candle to a wall, after giving them a box of tacks, candles, and matches. Most people attempted to complete the task by attaching the candle directly to the wall, either with a tack or by melting the candle. Few participants realized that the most effective solution utilized the box by attaching it to the wall with a tack, then resting the candle in the box. People "fixated" on the well-learned function of the box as a container, rendering them unable to conceptualize its alternative function as a shelf.

Problems with linguistic transfer and functional fixedness reflect a functional system; people adaptively default to the typical meaning of a word or use of an object rather than regularly reevaluate the range of a meaning or function, as this typically suffices. Reliance on a schema is similar; it works like heuristics in that it is cognitively efficient and generally leads one to the correct answer, but can also occasionally lead one astray. For example, when learning about complex phenomena, teachers and students often make comparisons to well-known objects and processes; sometimes these examples lead people astray when they overgeneralize the example. When computer science instructors discuss the concept of a *variable*, they often compare them to boxes—boxes hold things, just like a variable does. People possess a detailed schema for boxes (e.g., rectangular, often made of cardboard, holds things), and to some extent this helps. However, students may be misled by this analogy, as multiple objects fit in a box, but a variable possesses only one definition.

Larger problems arise when a learner brings along misconceptions or other erroneous beliefs. For example, many introductory psychology students possess preconceived notions about a wide range of course topics. Specifically, many believe that suicide rates peak in adolescence, that hypnosis can recover memories, and that polygraphs can accurately detect lies (Taylor & Kowalski, 2004). These misguided ideas set the stage for *proactive interference* (see Anderson & Neely, 1996 for a review), where students' misconceptions interfere with learning and remembering new, updated information.

4.6 Property #5: Partial Matches Are Often "Good Enough"

As described above, people often hold false beliefs or incorrectly apply otherwise correct knowledge; there are also situations in which people *neglect* their knowledge. That is, despite "knowing better," people miss referents that contradict their knowledge (i.e., *knowledge neglect;* see Marsh and Umanath 2013 for a review). Consider a passage about a plane that crashes on the border of France and Spain, which ends with *The authorities were trying to decide where to bury the survivors*. Readers generated a solution to this problem, although the question does not make sense, as survivors are alive and should not be buried. Despite knowing the definition of *survivors*, people often responded with answers like, "let the relatives decide" instead of noting that survivors should not be buried (Barton & Sanford, 1993). Similarly, in what is called the Moses Illusion, people willingly answer questions like *How many animals of each kind did Moses take on the ark?* despite later demonstrating knowing that the correct referent is *Noah* (Erickson & Mattson, 1981). The bottom line is that people often do not take advantage of the information stored in memory.

In line with Gilbert's ideas, it is adaptive to accept new information given that it is "close enough" to the correct answer, as opposed to carefully comparing incoming information to knowledge (*partial match theory;* Reder & Kusbit, 1991). This strategy makes sense when one considers that speech is full of disfluencies, including breaks, nonlexical utterances (e.g., "uh," "erm"), mispronunciations, and inappropriate word choices. Anyone who has ever sat through a recording of his own speech or presentation can attest to this; we have all probably thought, *do I really sound like that?* Comprehension presents an enormous challenge if people always algorithmically process language. Instead, we form shallow representations and use knowledge to fill in the gaps, allowing us to interpret garbled input streams like *in the grand scream of things* (*good enough theory;* Ferreira, Bailey, & Ferraro, 2002). Accepting speech that is "good enough" offers necessary flexibility, given that the same information often presents itself in slightly different forms.

Unfortunately, people accept close matches even when errors are not "slipped in" and unexpected; warnings are generally ineffective (e.g., Marsh & Fazio, 2006). Even trial-by-trial detection does not eliminate knowledge neglect. In one study, participants read stories sentence-by-sentence and (in one condition) made a keypress every time they detected an error. Those in the control condition received no special instructions other than a general warning that stories could contain errors. All participants read stories that included misleading references (e.g., *paddling around the largest ocean, the Atlantic*). Following the reading phase, participants answered general knowledge questions like *What is the largest ocean on Earth*?

Readers demonstrated some ability to detect errors: they were more likely to press the error key when sentences contained misinformation than when they contained correct or neutral references. However, this performance was hardly impressive; readers inappropriately flagged 20% of error-free statements, while also missing approximately two-thirds of the errors. Critically, Table 2 demonstrates that this pattern occurred regardless of participants' knowledge. Based on Nelson and Narens' (1980) norms, half of the facts were well-known (easy; answered on average by 70% of norming participants) and half were unknown (hard; answered by only 15% of norming participants). In both cases, participants caught approximately one-third of the errors (Marsh & Fazio, 2006). Additional experiments yielded similar results and suggested no advantage for detecting errors that contradicted known facts (e.g., Fazio & Marsh, 2008; Umanath & Marsh, 2012). Detection instructions do reduce the later reproduction of errors on a general knowledge test, but the reduction in suggestibility is relatively small as most errors still slip by the reader.

The assumption in the literature has been that knowledge neglect only occurs for weakly-held knowledge and that individuals would not fail to notice errors that contradict strong or expert knowledge. That is, "Biblically trained people [would] not false alarm to the Moses question" (Reder & Cleeremans, 1990, p. 249). However, there are reasons to question this prediction, given that experts rely on heuristics in many different tasks (e.g., Tversky & Kahneman, 1971). To test these ideas, we recruited biology and history graduate students to look for errors embedded in biology (e.g., *Water contains two atoms of helium and how many atoms of oxygen?*) and history questions (e.g., *In which state were the forty-niners searching for oil?*) (Cantor & Marsh, 2015). Overall, expertise helped: participants were more likely to detect errors that contradicted their expert knowledge. However, even experts missed approximately one-third of total errors, and they reproduced a small portion of these on a later general knowledge test. The experts did not appear to be "reading over" the errors, as **bolding and underlining** the key concepts did not help the experts any more than the non-experts. Instead, experts seemed to be relying on the same "good enough" strategy as the non-experts.

Table 2 *Error Detection During Story Reading*. Proportion of critical sentences labeled as containing an error as a function of question ease (easy or hard) and fact framing (misleading, correct, or neutral)

	Misleading	Correct	Neutral
Easy	.35	.17	.26
Hard	.31	.26	.17

Data are from Marsh and Fazio (2006) Experiment 3.

Of course, all of these examples include a referent that is "good enough," or semantically close enough to the truth for the partial match to be accepted. The Moses Illusion decreases when the referent is semantically or phonologically further from the target, such as when *Moses* is replaced with *Adam* (van Oostendorp & de Mul, 1990; see also Hinze, Slaten, Horton, Jenkins, & Rapp, 2014 for a related effect of plausibility). Monitoring takes effort, and accepting "good-enough" representations is a shortcut that normally works. Accordingly, use of this strategy increases in generally valid contexts, where monitoring is not worth the effort. For example, readers are less likely to catch a problem with *How many animals of each kind did Moses bring on the ark?* when tricky questions are rare (Bottoms, Eslick, & Marsh, 2010). Use of a "good-enough" approach also increases in attention-demanding contexts, given that few resources are left over for monitoring. For example, consumers of entertaining media, like novels and movies, who are busy building mental models to track multiple characters, storylines, and goals, can devote relatively few resources to catching factual inaccuracies.

5. LINGERING QUESTIONS ABOUT ERROR REPRESENTATION AND RETRIEVAL

5.1 Co-existence versus Overwriting

Importantly, we do not believe that errors overwrite or otherwise erase truths already stored in memory; instead, the literature suggests that both the error and correct knowledge can coexist in memory. The most telling evidence is that people can regain access to their correct knowledge, even after producing and using the misinformation. First, many knowledge checks occur *after* the main part of the experiment, and people access correct knowledge that they failed to leverage earlier (e.g., Bottoms et al., 2010). Similarly, the effects of misinformation dissipate over time; as the misinformation is forgotten, people recover access to their correct knowledge (Fazio et al., 2010). On the flip side, one can correct a misconception (see Section 6), but the misconception may re-emerge over time—again meaning that it must still be stored in memory (Butler, Fazio, & Marsh, 2011).

This claim is consistent with event memory, as argued within the eyewitness testimony literature, where the general conclusion was that the eyewitness' original memory was blocked rather than overwritten (McCloskey & Zaragoza, 1985). The effect of misinformation, whether about events or knowledge, is a simple example of retroactive interference, with a recently encountered error-blocking access to correct information stored in memory. The effects at retrieval are different, in that the eyewitness' problem is more likely to be one of source misattribution (Lindsay & Johnson, 1989), whereas the student's problem is that she interprets the ease with which misinformation comes to mind as evidence of truth. But in both cases, the representation is the same: two competing memories.

5.2 Direct Retrieval versus Construction

People often construct, rather than directly retrieve, truth. When one agrees with the statement that Lima is the capital of Portugal or that San Diego is larger than San Antonio, one is not necessarily retrieving that information directly from memory. Direct retrieval is often unnecessary given the shortcuts in the system for judging truth. As already reviewed, people have a bias to claim information is true, and evaluations may be based on how fluently something comes to mind rather than direct retrieval per se. It is clear that direct retrieval is not always attempted; people are very quick to say that they don't know the capital of Jupiter, for example—clearly no attempt at exhaustive search was made (Glucksberg & McCloskey, 1981).

The point here is that there is often an assumption in the literature that heuristics and constructions of knowledge only occur in cases of ignorance (e.g.,Unkelbach & Stahl, 2009)—but that is not the case. Having knowledge does *not* always preclude the use of these heuristics. We further unpack how this might happen in the next section.

5.3 A Fluency-Conditional Model of Illusory Truth

Intuitively, it seems unlikely that repeating *A date is a dried plum* makes people believe it. After all, most people know that drying plums produces *prunes*, not dates. Indeed, researchers assumed that illusory truth only emerges when statements are "ambiguous, that is, participants have to be uncertain about their truth status because otherwise the statements' truthfulness will be judged on the basis of their knowledge" (Dechêne et al., 2010, p. 239). In other words, people presumably only rely on fluency when they lack knowledge. Along these lines, Unkelbach and Stahl's (2009) model of illusory truth includes a knowledge parameter that is intentionally low; they used obscure materials, assuming that knowledge eliminates illusory truth.

Recent work in our lab, however, demonstrates the opposite: repetition can influence people's belief that *A date is a dried plum*. Participants read statements that contradicted well-known and unknown facts, as estimated by norms. After rating these and new statements' truthfulness, they completed a knowledge check to determine which specific facts they knew. In contrast to dominant assumptions, repetition swayed judgments of both known and unknown facts. This pattern emerged regardless of whether knowledge was estimated via norms or defined on the basis of individuals' knowledge check performance. In other words, fluency does sometimes "trump" knowledge (Fazio, Brashier, Payne, & Marsh, 2015).

The behavioral data are clear on their own, but we also tested these ideas using multinomial models. Figure 2 shows the two different models: the

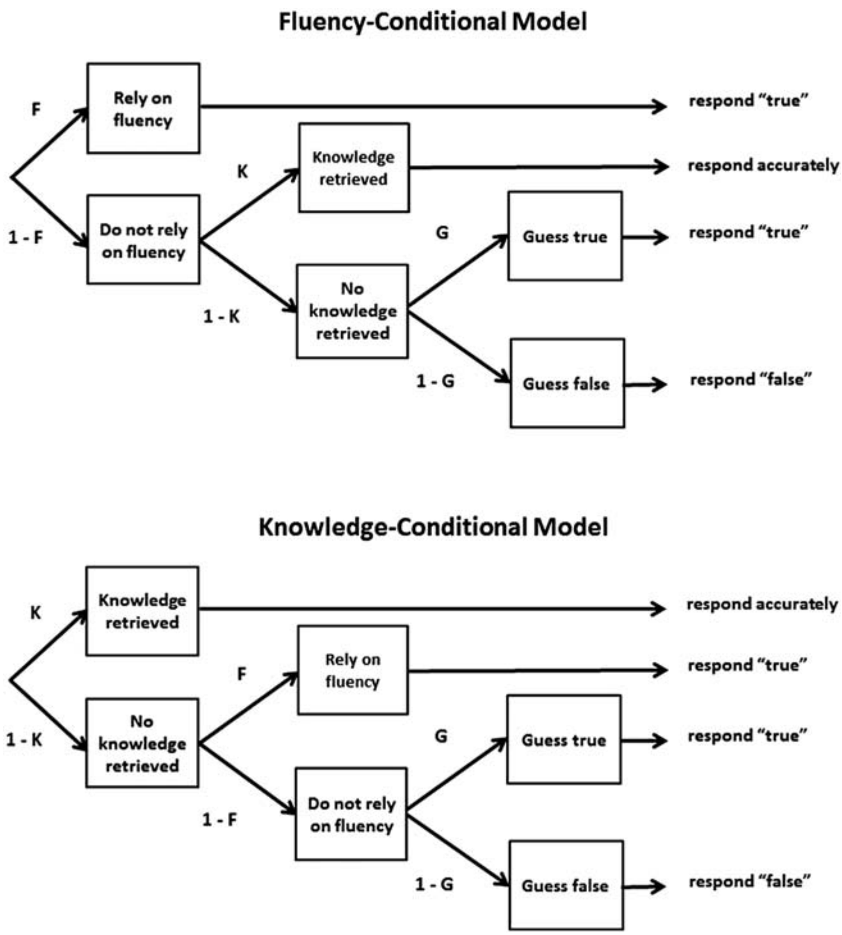

Figure 2 Fluency-conditional and knowledge-conditional models of illusory truth. K = knowledge; F = fluency; G = guess true. *Figure adapted from Fazio, et al. (2015).*

fluency-conditional model, wherein fluency can supersede knowledge, and the *knowledge-conditional model* assumed in the literature, where fluency only influences behavior in the absence of knowledge. In a new experiment requiring binary judgments, the behavioral result replicated and model testing revealed that the fluency-conditional model fit the data better than the knowledge-conditional model. In other words, people *sometimes* rely on fluency despite the fact that contradictory knowledge is stored in memory. Of course, people do not always rely on fluency over knowledge; fluency may be discounted (i.e., explained away) or absent due to inattention on a given trial. However, the success of the knowledge-conditional model demonstrates that it is possible to rely on this fluency signal, even when knowledge can be retrieved from memory. Extraneous factors, like repetition, can encourage people to accept contradictions of even well-learned facts.

6. CORRECTING ERRORS IN THE KNOWLEDGE BASE

6.1 Overview

Mistakes are not necessarily a bad thing. Much research shows the advantage of *desirable difficulties* (Schmidt & Bjork, 1992): challenges should be introduced into the learning phase, even if they cause frustration, slow the learning process, and lead to errors. More recently, Kornell and colleagues explicitly argued that trying to answer a question, and responding with an error, benefits memory, so long as feedback is provided. To avoid item-selection problems, they used fictitious stimuli that were *impossible* for participants to answer (borrowed from Berger et al., 1999, see Section 2.5) and weakly related word pairs (pond—frog). They compared the consequences of guessing and erring (followed by correct answer feedback) to the outcomes of simply studying the correct answers. Across experiments, forced guessing yielded performance that matched, and sometimes exceeded, studying (Kornell, Hays, & Bjork, 2009; see also Grimaldi & Karpicke, 2012; Kang et al., 2011). Why did error generation benefit learners? One possibility is that as people attempt an answer, they elaborate and think of related information (e.g., Carpenter, 2009), which facilitates the eventual encoding of the truth. Another possible explanation is that the error serves as a mediator (e.g., Pyc & Rawson, 2010) that connects the question and the correct answer (i.e., the error *green* could serve as an intermediary cue between *pond* and *frog*). Of course, learners needed to receive feedback to correct their answers; guesses unaccompanied by feedback can cause the same problems as the undetected errors discussed earlier.

6.2 Basic Advice for Correction

Correcting misinformation cannot take the form of simply pointing out the error (see Ecker, Swire, & Lewandowsky, 2013; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012 for reviews). Merely telling a learner that she is wrong leaves her with a gap in her knowledge, but nothing to fill it in with. Even if she knows that the information is wrong, she may continue to rely on it (i.e., *continued influence effect*). Studies illustrating this point involve participants reading stories (e.g., about a warehouse fire) and later learning that pieces of information from the story were false (e.g., the negligent storage of volatile materials) and should be disregarded (e.g., Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). Even though participants acknowledge that the retracted information is false, they still make inferences based on the misinformation (e.g., when asked the cause of the fire).

Similarly, research on feedback in educational contexts finds very limited benefits of telling students *whether* their test answers are correct or incorrect (*verification feedback*) (Pashler, Cepeda, Wixted, & Rohrer, 2005). Verification feedback works only when this feedback is delivered on a multiple-choice test (Marsh, Lozito, Umanath, Bjork, & Bjork, 2012). Knowing that one has chosen the wrong multiple-choice alternative allows students to narrow down the remaining possible answers. But when the test is composed of short-answer questions, rather than multiple-choice questions, verification feedback leaves students no closer to the correct answer.

Instead, the key to correcting errors is to replace the error with the correct answer (if there is one) or at least an alternative response. For example, in the story retraction example above, participants relied less on the false information if provided with an alternative account (e.g., that the fire was caused by petrol-soaked rags). Additionally, telling the learner the correct answer to a test question as feedback (e.g., via text, presentation, etc.) effectively decreases error reliance and increases correct performance on a follow-up test (e.g., Pashler et al., 2005). Surprisingly, this advice proves even more effective for correcting erroneous beliefs held with high confidence; when errors take the form of *Sydney is the capital of Australia* or *Water is stored in a camel's hump*, high confidence in the error predicts correction (*hypercorrection effect;* Butterfield & Metcalfe, 2001). This unexpected result has been replicated numerous times (e.g., Butler, Karpicke, & Roediger, 2008; Kulhavy, Yekovich, & Dyer, 1976), although the errors do start to re-emerge over time (Butler, et al., 2011). These results appear to reflect an attentional mechanism, whereby learners pay more attention to feedback when it surprises them (see Butterfield & Metcalfe, 2006 and Fazio & Marsh, 2009).

Unfortunately, sometimes there is no ideal alternative explanation. For example, while the general consensus is that vaccines do not cause autism (e.g., DeStefano & Thompson, 2004), the actual cause is still unknown. Without an alternative response to fill the gap left, such misconceptions are particularly challenging to combat (Ecker et al., 2013). Furthermore, sometimes the truth is known, but it is more complicated than the misconception. Because people tend to prefer simple explanations (Lombrozo, 2007), they may continue to fall back on a simple misconception if the truth is too complex.

A note of caution: when providing feedback, it is important to present the truth without actually repeating the misinformation (Ecker et al., 2013). While the intent may be to repeat the error to then provide the correction, repetition of errors also makes them more fluent and familiar. As a result, this attempt to correct misinformation can actually backfire and increase error reliance (e.g., Skurnik, Yoon, Park, & Schwarz, 2005). For example, correctly stating *The Big Dipper is not a constellation* repeats the association between the *Big Dipper* and *constellation*, potentially strengthening the misconception.

6.3 When Should Feedback Do More than Provide the Answer?

Sometimes exposure to the correct answer is not enough. For example, work from our lab finds that explanation feedback can be even better than correct answer feedback in certain situations (Butler, Godbole, & Marsh, 2013). To examine the benefits of explanation feedback, participants read passages about various topics (e.g., respiratory system, tropical cyclones) and took an initial short-answer test on concepts from the passages. After answering a question, participants either received correct answer feedback, explanation feedback, or no feedback. For example, for the question, *What is the process by which gas exchange occurs in the part of the human respiratory system called the alveoli*? correct answer feedback contained only the answer (*Gas exchange occurs within the alveoli through diffusion.*), while explanation feedback expanded on the answer by including two additional sentences that elaborated on the answer. (*Diffusion is the movement of particles from a region of high concentration to a region of low concentration. The oxygen concentration is high* *in the alveoli and the carbon dioxide concentration is high in the pulmonary capillaries, so the two gases diffuse across the alveolar membrane in opposite directions toward lower concentrations.*)

Two days later, participants returned to take a final test consisting of both repeated questions from the initial test as well as new inference questions that required participants to transfer their knowledge to a new context. Performance on repeated questions benefited from either kind of feedback relative to no feedback, with no difference between explanation and correct answer feedback. However, when the final test required transfer of learning, receiving an explanation improved performance above and beyond receiving the correct answer alone (Butler et al., 2013).

Similarly, research investigating strategies to invoke conceptual change finds that telling the learner the truth (e.g., via expository texts that only explain the truth) does not eliminate science misconceptions (see Guzzetti, Snyder, Glass, & Gamas, 1993 for a meta-analysis). Instead, strategies that induce "cognitive conflict" by directly refuting the misconception effectively promote conceptual change. For example, Hynd and Alvermann (1986) compared the effectiveness of standard texts that presented the correct information to ones that explicitly refuted the misconception while also presenting the correct information. The target was misconceptions about motion theory (i.e., about the path an object would take when launched off a cliff); a pretest confirmed that the majority of participants did hold a misconception about motion theory. Most students endorsed a false "impetus theory," where the projectile curves downward and then falls straight down. Next, students either read a standard text that explained Newtonian mechanics or a refutational text that directly contrasted impetus theory with Newtonian mechanics. On a final test administered 2 days later, students who read the refutational text changed their prior misconceptions more often than students who read the nonrefutational text.

6.4 Problem: Teacher and Student Preferences

One additional challenge is that the most effective feedback does not always match people's intuitions of what will help them the most. For example, many laboratory studies show that delaying feedback is a good thing, likely because it spreads out exposure to the information over time (spaced study; e.g., Butler & Roediger, 2008; Metcalfe, Kornell, & Finn, 2009). However, students do not appreciate this benefit. In our work, we delivered feedback either immediately after a homework deadline or 1 week later in a real engineering course (Signals and Systems) at the University of Texas El

Paso. Students thought they learned more with immediate feedback and disliked the delayed feedback—but the delayed feedback led to better performance on the final exam (Mullet, Butler, Verdin, von Borries, & Marsh, 2014).

A similar discrepancy emerges in students' preferences for the *source* of the feedback they receive. Students strongly prefer instructor feedback over peer feedback, even though peer feedback has the potential to match, or even exceed, the benefit of instructor feedback (e.g., Ertmer et al., 2007; Topping, 1998). We should note, of course, that not all peer feedback is high quality; successful peer feedback requires training and a careful scoring rubric (e.g., Falchikov & Goldfinch, 2000)—the point for present purposes is that it *can* be high quality. Regardless, students feel that instructors are unbiased, motivated to provide quality feedback, and more knowledgeable than other students (e.g., Ertmer et al., 2007). The teacher, on the other hand, may want to assign exercises for which there would be a large amount of time-consuming grading, or the instructor may view the act of providing feedback to be a valuable educational activity (e.g., Ertmer et al., 2007). This issue will continue to assert itself as coursework moves online and increasing numbers of students forces the instructor to choose between assignments that can be automatically graded (multiple-choice questions) versus using peer feedback to evaluate writing and solutions to problems. The Massive Open Online Courses (MOOCs) hosted by Coursera still draw thousands of students, and consequently depend upon peer feedback.

6.5 Problem: Misconceptions May Be Motivated

Misconceptions will be particularly sticky and hard to correct when they support someone's belief system and worldview (Lewandowsky et al., 2012). For example, one study showed that following a retraction, Republicans were less likely than Democrats to correct the misconception that weapons of mass destruction were found in Iraq (Nyhan & Reifler, 2010). Another study demonstrated that Americans were more likely than German participants to continue to rely on retracted information about the Iraq War (e.g., that weapons of mass destruction were found; Lewandowsky, Stritzke, Oberauer, & Morales, 2005). Furthermore, providing corrections that contradict an individual's worldview can even backfire and strengthen his or her belief in the misconception (e.g., Ecker, Lewandowsky, Fenton, & Martin, 2014; Nyhan & Reifler, 2010).

So how can these worldview-consistent misconceptions be corrected? At the moment, there is no perfect solution. However, Lewandowsky and colleagues (2012) offer two tactics that may help. First, to the extent that the correction can be framed within the individual's worldview, the correction will be more successful (e.g., Kahan, 2010). Second, people seem to be more receptive to correcting misconceptions after an opportunity for self-affirmation. For example, when participants first wrote about a value that was important to them and made them feel good about themselves, they were subsequently less likely to report misconceptions (Nyhan & Reifler, 2011).

7. CONCLUSIONS

7.1 The Science of Learning Is Not Different for Errors

Cognitive psychologists have uncovered many learning strategies that promote long-term retention of knowledge (see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013 for a review). For example, actively retrieving concepts leads to better memory than passively restudying them (e.g., Roediger & Karpicke, 2006), as does spacing out learning opportunities over time relative to massing them (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The literature focuses heavily on the learning of true information, but the same principles apply to errors. For example, retrieving errors increases their likelihood of being reproduced on later tests (Barber, Rajaram, & Marsh, 2008). We predict that other learning strategies that promote the learning of truths can promote the learning of falsehoods, such as spaced practice.

7.2 Comparing Errors of Event Memory to Illusions of Knowledge

We know a lot about errors of event memory, and how to encourage the misremembering of words, pictures, and even events from people's own lives. We know less, however, about how errors enter the knowledge base. In some ways, the two types of errors appear to be similar: both show interference effects, for example. Other properties of knowledge suggest that the two kinds of errors differ qualitatively from one another; for example, source-monitoring errors play a large role in errors of event memory (Johnson, Hashtroudi, & Lindsay, 1993), but are unlikely to produce "slips" of knowledge, which people often retrieve without reference to source.

7.3 Open Questions

Obviously many open questions remain, but we end by highlighting three with particularly important practical and theoretical implications. First, we still do not know enough about how to revise strongly held misconceptions such as *vaccines cause autism*. Such misconceptions pose practical challenges as they obviously cannot be recreated in the laboratory, and because the believers of such misconceptions do not want to be corrected. Second, little research speaks to individual differences and whether particular types of people may be especially prone to misconceptions. Third, from a theoretical perspective, we need to better understand the differences between illusions of knowledge and errors of event memories; knowledge differs from events in countless ways, including but not limited to the number of past occurrences, distance in the past, and richness of the original encounter. Future research should examine whether it is these factors, rather than a distinction between knowledge and events, that matter most in the learning and correction of errors.

REFERENCES

Abernathy, E. M. (1940). The effect of changed environmental conditions upon the results of college examinations. *The Journal of Psychology, 10*, 293—301.

Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science, 220*, 25—30.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork, & R. A. Bjork (Eds.), *Memory (Handbook of perception and cognition)* (2nd ed., pp. 237—313). San Diego, CA: Academic Press.

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making, 2*, 81—94.

Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: self-perceived expertise predicts claims of impossible knowledge. *Psychological Science, 26*(8), 1295—1303. Advanced online publication.

Bacon, F. T. (1979). Credibility of repeated statements: memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 241—252.

Bahrick, H. P. (1984). Semantic memory content in permastore: fifty years of memory for Spanish learning in school. *Journal of Experimental Psychology: General, 113*, 1—29.

Bahrick, H. P., Hall, L. K., & Baker, M. K. (2013). *Life-span maintenance of knowledge*. New York: Psychology Press.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 336—345.

Barber, S. J., Rajaram, S., & Marsh, E. J. (2008). Fact learning: how information accuracy, delay, and repeated testing change retention and retrieval experience. *Memory, 16*, 934—946.

Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge: Cambridge University.

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: shallow semantic processing and cohesion establishment. *Memory and Cognition, 21*, 477—487.

Bauer, P. J., & Jackson, F. L. (2015). Semantic elaboration: ERPs reveal rapid transition from novel to known. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 271—282.

Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: young children extend knowledge by integrating episodes. *Journal of Experimental Child Psychology, 107*, 452–465.
Bedard, J., & Chi, M. T. H. (1992). Expertise. *Current Directions in Psychological Science, 1*, 135–139.
Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*, 446–458.
Begg, I., & Armour, V. (1991). Repetition and the ring of truth: biasing comments. *Canadian Journal of Behavioural Science, 23*, 195–213.
Bellezza, F. S. (1984a). Reliability of retrieval from semantic memory: common categories. *Bulletin of the Psychonomic Society, 22*, 324–326.
Bellezza, F. S. (1984b). Reliability of retrieval from semantic memory: information about people. *Bulletin of the Psychonomic Society, 22*, 511–513.
Bellezza, F. S. (1984c). Reliability of retrieval from semantic memory: noun meanings. *Bulletin of the Psychonomic Society, 22*, 377–380.
Bellezza, F. S. (1987). Reliability of retrieving information from knowledge structures in memory: self information. *Bulletin of the Psychonomic Society, 25*, 407–410.
Bellezza, F. S. (1988). Reliability of retrieving information from knowledge structures in memory: scripts. *Bulletin of the Psychonomic Society, 26*, 11–14.
Berger, S. A., Hall, L. K., & Bahrick, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied, 5*, 438–447.
Bishara, A. J., & Lanzo, L. A. (2015). All of the above: when multiple correct response options enhance the testing effect. *Memory, 23*, 1013–1028.
Bonar, J., & Soloway, E. (1985). Preprogramming knowledge: a major source of misconceptions in novice programmers. *Human-Computer Interaction, 1*, 133–161.
Boster, J. S., & Johnson, J. C. (1989). Form or function: a comparison of expert and novice judgments of similarity among fish. *American Anthropologist, 91*, 866–889.
Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses illusion: failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory, 18*, 670–678.
Botwinick, J., & Storandt, M. (1980). Recall and recognition of old information in relation to age and sex. *Journal of Gerontology, 35*, 70–76.
Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177–220.
Brown, W. (1923). To what extent is memory measured by a single recall trial? *Journal of Experimental Psychology, 6*, 377–382.
Brown, A. S. (1988). Encountering misspellings and spelling performance: why wrong isn't right. *Journal of Educational Psychology, 80*, 488–494.
Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*, 204–223.
Brown, R., & McNeil, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.
Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1088–1100.
Butler, A. C., Dennis, N. A., & Marsh, E. J. (2012). Inferring facts from fiction: reading correct and incorrect information affects memory for related information. *Memory, 20*, 487–498.
Butler, A. C., Fazio, L. F., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high confidence errors return. *Psychonomic Bulletin and Review, 18*, 1238–1244.
Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105*, 290–298.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 918–928.

Butler, A. C., & Roediger, H. L., III (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory and Cognition, 36*, 604–616.

Butler, A. C., Zaromb, F. M., Lyle, K. B., & Roediger, H. L., III (2009). Using popular films to enhance classroom learning: the good, the bad, and the interesting. *Psychological Science, 20*, 1161–1168.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1491–1494.

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning, 1*, 69–84.

Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory and Cognition, 43*, 193–205.

Cantor, A. D., & Marsh, E. J. (2015). *In what US state were the forty-niners searching for oil: Sometimes history experts miss the error* (Manuscript in preparation).

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1563–1569.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chi, M. T., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Chi, M. T. H., & Koeske, D. R. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19*, 29–39.

Cho, H.-H., Kahle, J. B., & Nordland, F. H. (1985). An investigation of high school biology textbooks as sources of misconceptions and difficulties in genetics and some suggestions for teaching genetics. *Science Education, 69*, 707–719.

Choi, H., & Smith, S. M. (2005). Incubation and the resolution of tip-of-the-tongue states. *The Journal of General Psychology, 132*, 365–376.

Clancy, M. (2004). Misconceptions and attitudes that interfere with learning to program. In S. Fincher, & M. Petre (Eds.), *Computer science education research* (pp. 85–100).

Clement, J., Narode, R., & Rosnick, P. (1981). Intuitive misconceptions in algebra as a source of math anxiety. *Focus on Learning Problems in Mathematics, 3*, 36–45.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Conway, M. A., Cohen, G., & Stanhope, N. (1991). On the very long-term retention of knowledge acquired through formal education: twelve years of cognitive psychology. *Journal of Experimental Psychology: General, 120*, 395–409.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General, 126*, 393–413.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: a meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14*, 238–257.

DeStefano, F., & Thompson, W. W. (2004). MMR vaccine and autism: an update of the scientific evidence. *Expert Review of Vaccines, 3*, 19–22.

Dubeck, L. W., Bruce, M. H., Schmuckler, J. S., Moshier, S. E., & Boss, J. E. (1990). Science fiction aids science teaching. *The Physics Teacher, 28*, 316–318.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58*(5) (Serial No. 270).
Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4–58.
Ecker, U. K. H., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory and Cognition, 42*, 292–304.
Ecker, U. K. H., Swire, B., & Lewandowsky, S. (2013). Correcting misinformation – a challenge for education and cognitive science. In D. N. Rapp, & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. MIT Press.
Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: a semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*, 540–551.
Ertmer, P. A., Richardson, J. C., Belland, B., Camin, D., Connolly, P., Coulthard, G., et al. (2007). Using peer feedback to enhance the quality of student online postings: an exploratory study. *Journal of Computer-Mediated Communication, 12*, 412–433.
Eslick, A. N., Fazio, L. K., & Marsh, E. J. (2011). Ironic effects of drawing attention to story errors. *Memory, 19*, 184–191.
Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*, 287–322.
Farnsworth, P. R. (1934). Examinations in familiar and unfamiliar surroundings. *Journal of Social Psychology, 5*, 128–129.
Fazio, L. K., Agarwal, P. K., Marsh, E. J., & Roediger, H. L., III (2010). Memorial consequences of multiple-choice testing on immediate and delayed tests. *Memory and Cognition, 38*, 407–418.
Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993–1002.
Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin and Review, 15*, 180–185.
Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin and Review, 16*, 88–92.
Ferland, F., Ladouceur, R., & Vitaro, F. (2002). Prevention of problem gambling: modifying misconceptions and increasing knowledge. *Journal of Gambling Studies, 18*, 19–29.
Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.
Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107–119.
Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*, 601–613.
Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*, 221–233.
Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 311–325.
Gobbo, C., & Chi, M. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development, 1*, 221–237.
Goswick, A. E., Mullet, H. G., & Marsh, E. J. (2013). Suggestibility from stories: can production difficulties and source monitoring explain a developmental reversal? *Journal of Cognition and Development, 14*, 607–616.
Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory and Cognition, 40*, 505–513.

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: a comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly, 28*, 116–159.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208–216.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16*, 107–112.

Hawkins, S. A., & Hoch, S. J. (1992). Low-involvement learning: memory without evaluation. *Journal of Consume Research, 19*, 212–225.

Henkel, L. A., & Mattson, M. E. (2011). Reading is believing: the truth effect and source credibility. *Consciousness and Cognition, 20*, 1705–1721.

Hinze, S. R., Slaten, D. G., Horton, W. S., Jenkins, R., & Rapp, D. N. (2014). Pilgrims sailing the Titanic: plausibility effects on memory for misinformation. *Memory and Cognition, 42*, 305–324.

Hodges, J. R., & Patterson, K. (1995). Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia, 33*, 441–459.

Hynd, C., & Alvermann, D. E. (1986). The role of refutation text in overcoming difficulty with science concepts. *Journal of Reading, 29*, 440–446.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126*, 248–277.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: when misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1420–1436.

Kahan, D. (2010). Fixing the communications failure. *Nature, 463*, 296–297.

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology, 103*, 48–59.

Katsikopoulos, K. V. (2010). The less-is-more effect: predictions and tests. *Judgment and Decision Making, 5*, 244–257.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*, 1–24.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 989–998.

Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68*, 522–528.

Landau, J. D., & Bavaria, A. J. (2003). Does deliberate source monitoring reduce students' misconceptions about psychology? *Teaching of Psychology, 30*, 311–314.

Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science, 10*, 477–493.

Lee, G. M., Friedman, J. F., Ross-Degnan, D., Hibberd, P. L., & Goldmann, D. A. (2003). Misconceptions about colds and predictors of health service utilization. *Pediatrics, 111*, 231–236.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*, 106–131.

Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: the Iraq War 2003. *Psychological Science, 16*, 190–195.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory and Cognition, 17*, 349—358.
Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*, 232—257.
Marsh, E. J. (2004). Story stimuli for creating false beliefs about the world. *Behavior Research Methods, Instruments, and Computers, 36*, 650—655.
Marsh, E. J., Agarwal, P. K., & Roediger, H. L., III (2009). Memorial consequences of answering SAT II Questions. *Journal of Experimental Psychology: Applied, 15*, 1—11.
Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: difficulties in reducing reliance on fictional stories. *Memory and Cognition, 34*, 1140—1149.
Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. *Memory, 20*, 899—906.
Marsh, E. J., Lozito, J. P., Umanath, S., Bjork, E. L., & Bjork, R. A. (2012). Using verification feedback to correct errors on a multiple-choice test. *Memory, 20*, 645—653.
Marsh, E. J., Meade, M. L., & Roediger, H. L., III (2003). Learning facts from fiction. *Journal of Memory and Language, 49*, 519—536.
Marsh, E. J., & Umanath, S. (2013). Knowledge Neglect: failures to notice contradictions with stored knowledge. In D. N. Rapp, & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. MIT Press.
McCloskey, M. (1983). Intuitive physics. *Scientific American, 248*, 122—130.
McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114*, 1—16.
McNemar, A. (1969). *Psychological statistics* (4th ed.). New York: Wiley.
Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: do all roads lead to Rome? *Cognitive Psychology, 32*, 49—96.
Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory and Cognition, 37*, 1077—1087.
Metzger, R. L., Boschee, P. F., Haugen, T., & Schnobrich, B. L. (1979). The classroom as learning context: changing rooms affects performance. *Journal of Educational Psychology, 71*, 440—442.
Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227—234.
Mitchell, D. B. (1989). How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 31—49.
Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition, 3*, 222—229.
Munson, B. H. (1994). Ecological misconceptions. *The Journal of Environmental Education, 25*, 30—34.
Nakhleh, M. B. (1992). Why some students don't learn chemistry. *Journal of Chemistry Education, 69*, 191—196.
Nelson, T. O. (1988). Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals. In M. M. Gruneberg, P. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (Vol. 2, pp. 190—196). New York: Wiley.
Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19*, 338—368.
Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative photographs (or words) inflate truthiness. *Psychonomic Bulletin and Review, 19*, 969—974.

Newman, E. J., Garry, M., Unkelbach, C., Bernstein, D., Lindsay, D. S., & Nash, R. A. (2015). Truthiness and falsiness of trivia claims depend on judgmental contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(5), 1337–1348.

Nyhan, B., & Reifler, J. (2010). When corrections fail: the persistence of political misperceptions. *Political Behavior, 32*, 303–330.

Nyhan, B., & Reifler, J. (2011). *Opening the political mind? The effects of self-affirmation and graphical information on factual misperceptions* (Unpublished manuscript). Hanover, NH: Dartmouth College.

Odegard, T. N., & Koen, J. D. (2007). "None of the above" as a correct and incorrect alternative on a multiple-choice test: Implications for the testing effect. *Memory, 15*, 873–885.

van Oostendorp, H., & de Mul, S. (1990). Moses beats Adam: a semantic relatedness effect on a semantic illusion. *Acta Psychologica, 74*, 35–46.

Pachur, T., Bröder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: is recognition a non-compensatory cue? *Journal of Behavioral Decision-Making, 21*, 183–210.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3–8.

Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology, 14*, 330–345.

Phillips, D. L., & Clancy, K. J. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology, 77*, 921–940.

Potts, G. R., & Peterson, S. B. (1985). Incorporation versus compartmentalization in memory for discourse. *Journal of Memory and Language, 24*, 107–118.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science, 330*, 335.

Read, J. D., & Bruce, D. (1982). Longitudinal tracking of difficult memory retrievals. *Cognitive Psychology, 14*, 280–300.

Reber, R., & Schwartz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*, 338–342.

Reder, L. M., & Cleeremans, A. (1990). The role of partial matches in comprehension: the Moses illusion revisited. In A. Graesser, & G. Bower (Eds.), *The psychology of learning and motivation* (Vol. 25, pp. 233–258). New York: Academic Press.

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: imperfect encoding, retrieval, or match? *Journal of Memory and Language, 30*, 385–406.

Roediger, H. L., III., & Karpicke, J. D. (2006). The power of testing memory: basic research and implications of educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1155–1159.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science, 7*, 411–426.

Ross, P. E. (August 2006). The expert mind. *Scientific American*, 64–71. Retrieved from http://www.arts.uwaterloo.ca/~dkoehler/ACC784/Ross2006.pdf.

Saufley, W. H., Jr., Otaka, S. R., & Bavaresco, J. L. (1985). Context effects: classroom tests and context independence. *Memory and Cognition, 13*, 522–528.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31*, 713–724.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychonomic Bulletin and Review, 8*, 203–220.
Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: cognitive consequences of having information at our fingertips. *Science, 333*, 776–778.
Taylor, A. K., & Kowalski, P. (2004). Naïve psychological science: the prevalence, strength, and sources of misconceptions. *The Psychological Record, 54*, 15–25.
Thorndyke, P. W., & Hayes-Roth, B. (1979). The use of schemata in the acquisition and transfer of knowledge. *Cognitive Psychology, 11*, 82–106.
Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, 249–276.
Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1–12.
Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381–391.
Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110.
Umanath, S., Butler, A. C., & Marsh, E. J. (2012). Positive and negative effects of monitoring popular films for historical inaccuracies. *Applied Cognitive Psychology, 26*, 556–567.
Umanath, S., & Marsh, E. J. (2012). Aging and the memorial consequences of catching contradictions with prior knowledge. *Psychology and Aging, 27*, 1033.
Unkelbach, C. (2007). Reversing the truth effect: learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 219–230.
Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition, 18*, 22–38.
Van Der Wege, M., & Barry, L. A. (2008). Potential perils of changing environmental context on examination scores. *College Teaching, 56*, 173–176.
Van Overschelde, J. P., & Healy, A. F. (2001). Learning of nondomain facts in high- and low-knowledge domains. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1160–1171.
Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology, 40A*, 361–387.
Wynn, L. L., Foster, A. M., & Trussell, J. (2009). Can I get pregnant from oral sex? Sexual health misconceptions in e-mails to a reproductive health website. *Contraception, 79*, 91–97.
Yarmey, A. D. (1973). I recognize your face but I can't remember your name: further evidence on the tip-of-the-tongue phenomenon. *Memory and Cognition, 1*, 287–290.

CHAPTER FOUR

The Role of Stimulus Structure in Human Memory

Robert L. Greene[1]
Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, USA
[1]Corresponding author: E-mail: robert.greene@case.edu

Contents

Abstract

Long-term memory retention has long been seen as primarily or exclusively being driven by the meaning of stimuli. Three memory paradigms are reviewed that demonstrate the importance of stimulus structure in memory. First, memory is influenced by orthographic distinctiveness. Stimuli that have unusual spelling patterns are more likely to be processed as individual stimuli than are items that are less distinct. This may either enhance or degrade memory performance, depending upon the nature of the test that is given. Second, the phenomenon of recognition without identification shows that orthographic features can be effectively used as retrieval cues even when the attempted retrieval of a stimulus' name and meaning ends in failure, preventing use of semantic cues. Third, the pseudoword effect, the higher rate of positive responses to pronounceable nonwords than to words, is a demonstration that adding meaning to orthographic features may lead to a reduction of familiarity. Structural features can play an important role in long-term retention.

1. INTRODUCTION

Allport (1937) first brought the distinction between idiographic and nomothetic science to the field of psychology. Idiographic approaches

Psychology of Learning and Motivation, Volume 64
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2015.09.004

emphasize the unique, singular, and subjective aspects of experience. Nomothetic approaches emphasize the search for universal laws that can bring order and simplicity to a vast field of findings. This distinction is closely related to Cronbach's (1957) discussion of the two disciplines of scientific psychology.

Ever since the time of Ebbinghaus (1885/1964), the psychology of memory has been overwhelmingly nomothetic, utilizing list-learning experiments to draw empirical conclusions about the behavior of populations of participants. Early psychological researchers seemingly believed that their ultimate goal was to discover laws of learning. Thorndike (1911) stated that the psychology of learning can be reduced to two principles, the law of effect and the law of exercise. A number of laws are presented in the first standard textbook of verbal learning (McGeoch, 1942), showing the eagerness with which early memory researchers proclaimed universal laws. Curiously, however, although this nomothetic tendency has persisted in the psychology of memory even as behaviorist influence waned, theorists during the cognitive era have shown a surprising reluctance to specify the universal laws being demonstrated in our experiments. Cohen (1985) noted that memory textbooks rarely present laws of memory; even a book entitled *Principles of Learning and Memory* (Crowder, 1976) somehow neglected to mention any principles. This might be part of a larger trend in psychology in recent decades in which theorists seem to avoid presenting general laws of behavior or mental processes (Teigen, 2002); a field that once delighted in proclaiming universal laws of behavior now hesitates to draw any broad conclusions at all and falls back on narrowly specified explanations (Beach, 2014). We have become a field of exceptions without rules.

One principle that has been articulated and tremendously influential is that of levels of processing (Craik & Lockhart, 1972). Presented not as a theory of memory but rather as a framework for conceptualizing previous findings and guiding future research, this approach emphasizes that memory for an event is determined by the coding processes carried out on that event. Such coding events can be ordered from the most structural and perceptual levels at one end to semantic and pragmatic levels at the other. Retention is determined by the depth to which an event had been processed. Structural information is retained only briefly, whereas meaningful processing persists indefinitely and can underlie long-term memory performance. Indeed, classic experiments carried out by Craik and Tulving (1975) demonstrated that manipulations of processing depth can lead to dramatically different rates of memory performance. The challenge that Craik and Tulving faced was that processing is largely under the control of experimental participants. They addressed this challenge by using speeded orienting tasks and incidental learning. When orienting tasks encouraged structural processing of stimuli, later incidental memory was quite poor. When semantic processing was required by the orienting tasks, later retention was excellent.

The levels-of-processing framework laid out the importance of meaningful processing in an explicit form. However, this was not a particularly controversial aspect of this framework, as theoretical approaches developed both before and after Craik and Lockhart (1972) took similar views. Stage models of memory (e.g., Atkinson & Shiffrin, 1968; Waugh & Norman, 1965) typically distinguished between short-term and long-term memory stores in part on the nature of the encoded trace; short-term traces were sensory and phonological, while long-term traces were based on semantic features. More recent approaches, such as the distinction between data-driven and conceptually driven memory processes (e.g., Roediger, 1990; Roediger, Weldon, & Challis, 1989) saw standard explicit memory tests as being primarily conceptually driven. The empirical definition of conceptually driven is that performance on a memory task is enhanced by stimulus generation and by semantic orienting tasks, and recall and recognition would both clearly be conceptually driven by this definition.

Essentially, psychologists assumed that structural features can be important for short-term retention but that long-term retention is based entirely on meaning. This assumption can seem rather plausible. As cognitive psychologists first focused on short-term retention, one of the most striking patterns is its seeming reliance on phonological coding: the errors people made in immediate memory were largely based on sound, not meaning (Conrad, 1964), and performance was often strikingly poor when remembering lists of phonologically similar items in order (Conrad & Hull, 1964). Immediate retention is greatly affected by whether the input modality was auditory or visual (Crowder & Morton, 1969) or whether visually presented material was read aloud or silently (Conrad & Hull, 1968; Crowder, 1970). Examination of serial-position curves in free recall showed that variables related to meaningful processing could have large effects on early items, presumably retrieved from long-term storage, while having no effect on recall of items from the last few positions, presumably reflecting short-term storage; such variables include word frequency (Raymond, 1969; Sumby, 1963), semantic similarity of the stimuli on a list (Craik & Levy, 1970; Glanzer & Schwartz, 1971), or semantic orienting tasks (Glanzer & Koppenaal, 1977;

Seamon & Murray, 1976). Sensory variables affected only recall of the recency segment of lists (Conrad & Hull, 1968; Crowder, 1970; Crowder & Morton, 1969). The reflexive equating of short-term storage with structural features and long-term storage with semantic features had critics early on (e.g., Crowder, 1982; Shulman, 1971) but met with general acceptance as at least an approximation to the truth, as shown by its inclusion even in theoretical frameworks, such as levels of processing (Craik & Lockhart, 1972), that do not explicitly require multiple stores.

Ever since the original presentation of the levels-of-processing framework, memory researchers have emphasized the importance of meaningful processing. This emphasis has proven so useful in explaining empirical findings that Roediger (1993, p. 511) nominated the levels-of-processing principle as "cognitive psychology's most successful theory of learning and memory"; when he later reviewed laws of memory, levels of processing was discussed first and at the greatest length (Roediger, 2008). Its success as a conceptual framework for a vast amount of memory research is unquestioned. However, it has become increasingly clear over the years that structural aspects of stimuli are retained neither as briefly as Craik and Lockhart (1972) initially envisioned nor as inconsequential for long-term memory performance as most of us at one time believed.

It must be acknowledged that the concept of memory has expanded greatly since the days of two-store theories and levels of processing. In part, this expansion has brought into memory research topics that earlier theorists never intended to address and may not ever have been expected to benefit from meaningfulness. For example, memory for physical activities has now been systematically studied, but, given the motoric nature of this material, few would ever have been surprised that semantic processing will not always have a beneficial effect (Cohen, 1985; Engelkamp & Dehn, 2000). Similarly, data-driven repetition priming tasks have been carefully investigated and do not necessarily show levels-of-processing effects (e.g., Jacoby & Dallas, 1981; Roediger, Stadler, Weldon, & Riegler, 1992), though such effects are actually found in some designs (Challis & Brodbeck, 1992; Thapar & Greene, 1994); such priming tasks are inherently perceptual in nature and can easily be distinguished from more standard memory tasks. Roediger (2008) noted that a broadened definition of memory challenges many of our most common empirical generalizations about retention. Still, even when our focus is restricted to the standard list-learning paradigms of verbal learning that constitute the core of the memory literature that Craik and Lockhart (1972) intended to explain, structural features may be far more important in long-term retention than earlier theorists imagined.

It should also be acknowledged that it was clear decades ago that some aspects of perceptual experience must persist in memory; after all, recognition of faces or voices is likely to be impossible without such persistence. Participants are often able to remember the modality (auditory or visual) in which particular words are presented in mixed-modality lists (e.g., Hintzman, Block, & Summers, 1973), and, at least under narrowly defined circumstances, long-term retention may be affected by presentation modality (Gardiner & Gregg, 1979; Glenberg, 1984; Greene, 1985). False memory is affected by the modality of the presented lists (Cleary & Greene, 2002; Smith & Hunt, 1998). It has long been known that long-term memory can be affected by whether participants read visually presented words aloud or silently (Greene & Pearlman, 1996; Hopkins & Edwards, 1972), though it is only in recent years that this pattern received both a name (the production effect) and a sustained investigation (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). Clearly, the generalization that the perceptual details involved in the presentation of a list item are quickly lost is far more limited than we originally believed, even if the effects of such details are often marginal on tests of long-term retention. Still, this episodic perceptual information resulting from the specific details of list presentation is distinguishable in principle from intrinsic structural properties of stimuli included in lists, such as the orthographic features of words. The extent to which intrinsic structural aspects of stimuli, unrelated to the perceptual details of presentation, influence long-term memory will be the focus here.

Three separate literature are reviewed here, each demonstrating the role that the structural features of list items may play in long-term retention. The first literature reviewed deals with the effects of orthographic distinctiveness on memory. Stimuli with unusual orthographic features will be processed differently than those with more common orthographies; these differences in processing may lead to either better or worse memory performance, depending on the nature of the test that is given to participants. Second, the literature on recognition without identification will be reviewed in support of the proposition that orthographic features may be effective retrieval cues even in the absence of semantic detail. Third, the pseudoword effect will be discussed, a demonstration that familiarity may be enhanced if recognition is based only on orthographic features and that the addition of meaning may actually reduce the experience of familiarity.

2. EMPIRICAL DEMONSTRATIONS OF STRUCTURAL EFFECTS ON MEMORY

2.1 Effects of Orthographic Distinctiveness

Consider words such as "lymph," "khaki," and "Afghan." These words are no less meaningful than words such as "leaky," "kennel," and "airway." However, when native English speakers are asked to rate words for visual weirdness (however that may be interpreted by the people doing the rating), the first set is rated as much weirder than the second set. These are examples from Hunt and Elliott (1980), who used visual-weirdness ratings to define sets of orthographically common and distinct words. When participants were given a list composed of equal numbers of orthographically common and distinct words and were subsequently asked for free recall, an orthographic-distinctiveness effect was found: recall was much higher for the distinct words than for the common words. This effect was present only when the sets of words were randomly intermixed. No difference between the sets was found when they were presented on separate lists, suggesting that some sort of contrast is needed for visual weirdness to affect long-term retention.

Choosing sets on the basis of ratings of visual weirdness is one way to vary orthographic distinctiveness, although it can be unclear exactly how raters are interpreting the instructions and which aspects of visual structure are being utilized. An alternative way to examine the effects of orthographic distinctiveness is to select stimuli on the basis of an objective metric. Following the example of Cortese, Watson, Wang, and Fugett (2004), Glanc and Greene (2007, Experiment 1) examined the effects of orthographic neighborhood size on recognition memory. Orthographic neighborhood size was defined as the number of words that could be created by changing one letter of a target word while keeping other letters in their original position and minimizing phonological change (see Cortese et al., 2004). High-N words, such as "boy," "duck," and "bunch," were those with many similar words, where low-N words, such as "rev," "bulk," and "width," were relatively distinct. When subjects were shown a list composed of low-N and high-N words and were subsequently given a yes–no recognition test, an effect of orthographic distinctiveness was found, with low-N words receiving both more hits and fewer false alarms than high-N words.

A pattern such as that found with orthographic distinctiveness, where a stimulus class that receives more hits in recognition will have fewer false alarms than another stimulus class, is known as a mirror effect (Glanzer & Adams, 1985). Another simple way of demonstrating mirror effects in recognition is through null pairs in forced-choice testing. Typically, on a forced-choice test, the participant is given test pairs and asked to pick the member of each pair that had occurred on the list. However, on null pairs, there are no correct answers. A null pair consists either of two old items or two new items. Participants are not told the nature of the test pairs but are rather led to believe that one and only one member of each test pair had occurred on the study list; debriefing after the experiment typically shows that few subjects catch on to the tricky design. Because null pairs control for strength differences as a result of list presentation, they are a particularly powerful way to test for differences between stimulus classes. When a mirror effect is found, this means that the class of stimuli that seems particularly old when old also seems particularly new when new. Such a mirror pattern was found when neighborhood size was manipulated on null pairs that all contained one low-N word and one high-N word (Glanc & Greene, 2007, Experiment 2). When both words in a null test pair were old, participants tended to pick the low-N option as the stimulus they believed to be old. When both words in a null test pair were new, participants tended to pick the high-N option as the stimulus they believed to be old. Old low-N words are seen as "more old" than old high-N words, and new low-N words are seen as "more new" than new high-N words. This is the classic mirror-effect pattern.

At least subjectively, low-N words appeared to be more likely to be recollected consciously. The subjective experience of recognition is often studied through the use of remember/know judgments. On a recognition test using the remember/know procedure, after every item that receives a positive recognition response, the participants are asked to indicate whether they specifically remember the experience and detail of that item's occurrence or whether they merely know it was on the list through a general feeling of familiarity. When subjects were asked whether they specifically remembered the occurrence of the test words or merely knew that they had been on the list, the low-N advantage was found only on words identified as being remembered; high-N words were more likely to be classified as merely known than low-N words (Glanc & Greene, 2007, Experiment 3). The correct interpretation of these "Remember" and "Know" responses in recognition memory has been a highly controversial topic (e.g., McCabe, Geraci, Boman, Sensenig, & Rhodes, 2011; Mickes, Searle-Carlisle, & Wixted, 2013; Selmeczy & Dobbins, 2014; Wixted & Stretch, 2004; Yonelinas & Jacoby, 2012) and should be currently viewed as an unsettled issue; still, it is clear that, at least from the phenomenological perspective of the research participants, low-N words were more likely to lead to correct impressions of recollection than high-N words.

Word frequency is the variable most often studied with respect to mirror effects, with low-frequency words having more hits and fewer false alarms than high-frequency words (Glanzer & Adams, 1985). Malmberg, Steyvers, Stephens, and Shiffrin (2002) had speculated that the word-frequency effect in recognition is actually a result of orthographic distinctiveness, as relatively rare words are more likely to have unusual combinations of letters than common words. The finding by Glanc and Greene (2007) that orthographic distinctiveness, when manipulated directly, leads to a mirror pattern similar to that found with word frequency, is consistent with this claim. Orthographic distinctiveness may also underlie the word-length effect, as researchers have often confounded the length of words with the number of unique letter patterns (Jalbert, Neath, Bireta, & Surprenant, 2011). Orthographic distinctiveness may play a much greater role in long-term retention than we previously imagined.

Glanc and Greene (2007) speculated that the advantage for low-N words in recognition reflected differences in encoding between the stimulus classes. When participants study a list of words for later recall, they can focus on the properties of the individual items in isolation (i.e., item encoding) or they can try to relate stimuli to each other (i.e., relational encoding). The unusual physical appearance of orthographically distinct words may lead participants to emphasize item encoding on these items. This enhanced item processing of low-N words can lead to improved recognition performance. However, this may be reflected in worse performance on other tests of memory, which place less emphasis on pure item encoding. Glanc and Greene (2009) examined the effects of orthographic distinctiveness on associative recognition. In an associative-recognition test, participants study a list of pairs of stimuli. At the time of test, they are again given pairs. Some of the pairs are intact, containing stimuli that had been presented together on the memory list. Other test pairs are rearranged, containing words that had been included on the study list but not with each other. Participants have to discriminate between intact and rearranged pairs, giving positive responses to words that had been studied together and negative responses to words that had been studied separately. Because all of the individual words had been presented on the study list, item encoding processes are of no use on a task like that. Rather, accuracy in associative recognition is a reasonably pure measure of relational processing, the extent to which the two stimuli had been associated with each other during study. Glanc and Greene (2009) tested associative recognition for pairs of low-N words and pairs of high-N words. They found that associative recognition was affected by orthographic distinctiveness in a pattern that was opposite to that found with item recognition: in associative recognition, pairs of high-N words received both more hits and fewer false alarms than pairs of low-N words, a demonstration of a mirror pattern in associative recognition (cf. Greene, 1996).

An emphasis on processing trade-offs does not necessarily lead to clear predictions about free recall, a complicated task where item, relational, and order processing all contribute, with the exact mix often depending on complex contextual issues. Still, Cortese et al. (2004; see also Cortese, Watson, Khanna, & McCallion, 2006) examined orthographic neighborhood effects in free recall and found a low-N advantage. Presumably, in these experiments, the enhanced item encoding of low-N words outweighed whatever advantage in relational processing may have applied to high-N words.

McDaniel, Cahill, Bugg, and Meadow (2011) argued that an item-order account could be used to explain the effects of orthographic distinctiveness on memory. That is, they assumed that the enhanced item processing that is devoted to distinct items could lead to reduced memory for order. The extra time that participants spend processing item characteristics takes away from processing useful in situating items at particular locations on the list. McDaniel et al. argued that orthographic-distinctiveness effects are part of a more general pattern, where item information and order information may be systematically traded off against each other. Indeed, accounts of this sort were first developed to explain generation effects (Nairne, Riegler, & Serra, 1991) and have been broadly applied. However, the empirical pattern is a bit more complex than this. Memory for order can be based on several different types of information, which can be differentially affected by item and relational processing (Crowder & Greene, 2000). Indeed even in the paradigmatic case of generation, the effects found depend on the precise test of order information that is given (Greene, Thapar, & Westerman, 1998).

Remembering when a particular stimulus occurred can be tested in several different ways, and each method may require a unique combination of processes. Clearly the oldest method is serial learning, with a history that reaches into the start of experimental psychology. Ever since the time of Ebbinghaus (1885/1964), it has been argued that interitem associations may play a critical role in the recall of lists in order; although strict reliance on the chaining of item-to-item reliance as the sole basis of order information has been ruled out for decades, some role for interitem associations is hard to dispute (Crowder & Greene, 2000). Given the role that interitem associations may play in this task, stimuli that promote relational processing should enhance performance. Glanc and Greene (2012) asked participants to recall lists of six words in order; performance was higher when the list was composed of high-N words than when it was composed of low-N words. However, other subjects were given an order-reconstruction task. In this test, after presentation of a list of six items, the six items were shown on the screen in random order, and participants had to type in a number for each word to indicate the serial position that it had occupied on the preceding list. Performance was better when the list contained only low-N words than when it contained only high-N words. Possibly a critical element of the procedure was that participants were required to type the position numbers of test items presented in random order. The random order of testing presumably discouraged a reliance on serial item-to-item associations. Jalbert et al. (2011, Experiment 2) used a related procedure, a strict serial reconstruction task in which participants had to recreate the original list ordering of words by responding to items in the order that they had been studied. They found a high-N advantage on this serial reconstruction task, suggesting that serial ordering of responses is crucial for finding positive effects of large orthographic neighborhoods on order information. Even slight changes in procedure seem to modify the processes through which people remember order information.

Orthographic distinctiveness can have powerful effects on memory, but the direction of this effect will depend on the sort of information required by the test. When item processing is appropriate for a test, stimuli that are orthographically distinct (such as words with low neighborhood sizes) are advantaged. When relational processing is appropriate, an advantage for words with less distinct orthography (such as words with high neighborhood sizes) can be found. The encoding context of stimuli may also be critical. Kang, Balota, and Yap (2009, Experiment 2) approached this issue from the perspective of dual-route models of reading (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), which assume that reading printed words aloud may be done either by a nonlexical route in which phonology is created through the use of spelling-to-sound correspondence rules or by a lexical route in which the letter string is used to access a lexical entry, allowing retrieval of pronunciation information. Kang et al. argued that a recognition advantage for low-N words would be found only if participants rely on the nonlexical route during encoding of the words; that is, orthographic structure matters only if attention is focused directly on orthographic features during encoding. Kang et al. found results consistent with their predictions, in which they encouraged lexical-route identification or nonlexical-route identification by embedding the critical low- and high-N words in a series of visual stimuli that had to be pronounced as rapidly as possible. The visual stimuli were either nonwords (the occurrence of which would presumably prime reliance on nonlexical processes for adjacent items) or words (which would prime lexical processing of adjacent items). The priming of lexical processing eliminated the effects of orthographic distinctiveness on later recognition. For our purposes, the critical finding is that an encoding manipulation could either enhance or eliminate the low-N advantage in recognition memory, a result consistent with the claim that the effects of orthographic distinctiveness result from encoding differences between low-N and high-N words.

Orthographic distinctiveness influences long-term retention. The empirical picture is consistent with the hypothesis that item processing is enhanced for physically distinctive words. This can lead to an advantage for low-N words in item recognition (Glanc & Greene, 2007), which largely reflects item processing, and an advantage for high-N words in associative recognition (Glanc & Greene, 2009), which is sensitive to relational processing. Long-term retention is sensitive to the orthographic structure of the material, but the direction of the effect will depend on the nature of the information being tested.

2.2 Structural Cues at Retrieval: Recognition without Identification

Memory researchers often like to attribute particular effects to either encoding or retrieval. However, there is a curious asymmetry in these attributions. Attributions to encoding can never be made confidently because any changes happening to encoding may influence the later retrieval process. Discussion of orthographic distinctiveness effects typically emphasizes encoding processes because a reasonably coherent story can be told that way. However, clearly orthographic distinctiveness has consequences for retrieval, so no one is ever able to label an effect such as this an encoding effect.

Attributions to retrieval, however, can usually be made with more confidence. Because retrieval occurs only after encoding is completed, a manipulation that is applied only at that time can be assigned to the retrieval stage. Thus, one can try to directly study whether structural features of words are useful in retrieval.

One method to do this is recognition without identification. Peynircioglu (1990) first demonstrated this phenomenon. Participants are shown a list of words. Following the study list, she presented a test consisting of four-letter fragments, with each fragment uniquely identifying one word in the English language. Participants first have to attempt to complete each fragment with the appropriate word. Then, whether or not they were successful at completing the fragment, they had to give a rating of the likelihood that the solution word had been included on the study list. When a fragment had been successfully completed, this is merely a standard recognition test. However, even when participants were unable to complete a fragment, they still gave higher recognition ratings when the solution was a studied word than when it was not. That is, even when people could not identify a word, they nevertheless showed some degree of episodic memory for whether it had been on a studied list. If participants are unable to identify the word, they presumably are not able to access whatever meaningful processing was carried out on the word during encoding. Recognition would have to be based on the individual letters, even though the studied stimuli had been processed as complete words during encoding.

Cleary and Greene (2000) followed up on Peynircioglu's (1990) original finding, both replicating the effect even when procedural safeguards were added to eliminate relatively uninteresting response-bias explanations and extending it to new situations. Participants showed above-chance recognition memory even when recognition was based on unique two-letter fragments that they could not complete. The structural information underlying the effect does not seem to be visual or perceptual, as recognition without identification was found when the study list was presented in a different case or in a different sensory modality (i.e., auditory presentation of the study list) than the visually presented word fragments employed on the test.

Cleary and Greene (2001) extended the finding of recognition without identification to judgments of presentation frequency, a task that is also presumed to depend largely on evaluation of stimulus familiarity (Greene, 2008). However, no comparable effect was found when participants were tested on either associative recognition or list discrimination, two tests that require conscious recollection of the original encoded experience. This suggests that recognition without identification reflects familiarity processes, driven by a matching process of abstract representations of structural features of the originally encoded list words and the tested fragments. Ryals and Cleary (2012) provided additional evidence that familiarity driven by feature matching underlies these recognition phenomena.

One condition tested by Cleary and Greene (2001) is perhaps worth particular mention. Here, the standard order of information is reversed. Participants are given a study list containing four-letter fragments that could uniquely identify an eight-letter word. They try to solve the word, though they rarely succeeded. After the study list had been shown, a test is given. The test consists only of words, half of which corresponded to fragments that had been presented on the study list and half of which did not. The task is to give positive recognition responses to words that corresponded to studied fragments. For fragments that had been unsolved on the study list, meaningful processing would have to be minimal. Still, participants showed recognition memory well above chance even when the tested word corresponded to an unsolved fragment. This is recognition memory based purely upon orthographic features.

Cleary and Greene (2004) showed that this sort of feature-based recognition was not dependent on the use of word fragments at test. They employed a perceptual-identification test, where words are flashed quickly so that participants are able to identify some, but not all, of them. A list of words was studied. Then, in the test phase, words were flashed quickly, and participants had to try to identify each one. Whether or not identification was successful, ratings had to be made as to the likelihood that the flashed stimulus had been included on the study list. Even when they could not identify a flashed word, participants still showed recognition performance well above chance. Whatever visual features sampled during the perceptual-identification test were evidently sufficient to support recognition even if they could not be used to identify the word. This phenomenon of recognition without perceptual identification appears to be based upon familiarity, as no such effect was found on recollection-based tests, such as associative recognition (Cleary & Greene, 2005).

It would be easy enough to dismiss these various demonstrations of recognition given retrieval failure as cute laboratory demonstrations without any ecological validity. Still, they support a conclusion that memory based on structural features can be surprisingly powerful. It makes little sense to describe this recognition as being based on meaning or conceptually driven processes; being able to know that R _ _ _ _ _ _ P represents a studied word even when you cannot complete the fragment as RAINDROP is a compelling demonstration of structurally based recognition. Moreover, this is something that may occur at times outside the laboratory. Cleary (2014) argues that recognition unaccompanied by successful retrieval underlies common phenomena such as déjà vu and the tip-of-the-tongue effect.

The importance of cueing in successful retrieval has long been generally accepted. The phenomenon of recognition without identification has demonstrated that the range of effective memory cues may be broader than we originally imagined. A handful of letters in an unsolved letter fragment or a set of visual features in an unidentified stimulus can form the basis of accurate recognition.

2.3 Structurally Based Familiarity: The Pseudoword Effect

Word-frequency effects have been one of the most heavily studied topics in the study of recognition memory. Although it is not clear why this is so, one possible reason is that the contrast between the low-frequency advantage in recognition and the high-frequency advantage in recall provides perhaps the most compelling example we have of testing situations leading to different effects. A second reason is that word frequency has provided the strongest and most convincing evidence for the mirror effect (Glanzer & Adams, 1985), the finding that, in recognition studies involving two distinguishable classes of stimuli, the class with the higher hit rate often has the lower false-alarm rate so that there is no difference in overall numbers of positive responses between the two classes. The mirror effect has been portrayed as one of the fundamental regularities of recognition memory (Glanzer, Adams, Iverson, & Kim, 1993) and (along with the list-strength effect) as one of the two phenomena "that are at the heart of testing and evaluating the models" of recognition memory (Ratcliff & McKoon, 2000, p. 575). However, it should be noted that it is now clear that many exceptions to this mirror pattern may occur. Indeed, this is a law of memory for which there may be as many exceptions as there are examples (Greene, 2007; Roediger, 2008).

Most of these word-frequency studies simply compared high-frequency words with low-frequency words. However, a handful of studies included very low-frequency words, that is, words that are considered to be part of the lexicon but rare enough that they are unlikely to be known by any of the participants. Alternatively, real pseudowords (that is, nonwords that follow all of the orthographic norms of English) could be included instead. A condition with very low-frequency words or pseudowords (from here on referred to simply as pseudowords) was typically not the focus of theoretical interest. Still, when experiments containing this condition are reviewed, striking empirical consistencies can be found. First, participants' recognition memory accuracy (whether measured by the difference between hits and false alarms or by signal-detection measures) was much worse for pseudowords than for either high- or low-frequency words. This is consistent with the traditional emphasis on meaningfulness as a determinant of recognition memory. Second, participants gave positive recognition responses far more often to pseudowords than to genuine words (see Hintzman & Curran, 1997; Hockley & Niewiadomski, 2001; Whittlesea & Williams, 2000, 2001; Wixted, 1992 for examples of both of these empirical patterns). Greene (2004) called the greater number of positive recognition responses to pseudowords (or extremely low-frequency words) than to known words the *pseudoword effect*. In early papers, this pseudoword effect was not the focus of sustained investigation, but typically some attempt was made to account for this unexpected (but recurring) finding. The explanations researchers offered for this effect were typically variations of overcompensation: that is, participants knew that pseudowords would be less memorable than words and therefore adjusted their recognition process in some way to take this into account; however, they systematically overcompensated so that positive recognition responses became much more likely to pseudowords than to words. This overcompensation could be done by either shifting the criterion too far for pseudowords (Hockley & Niewiadomski, 2001; Stretch & Wixted, 1998) or by rescaling the familiarity values (Hintzman & Curran, 1997) or fluency values (Whittlesea & Williams, 2000, 2001) for pseudowords.

The memory literature is filled with demonstrations of the errors people make. Still, it is startling to see theorists assume that these errors with pseudowords all go in one direction (overcompensation), especially in the absence of any evidence that people are trying to compensate for anything and of any arguments as to why overcompensation should be more common than undercompensation. Greene (2004) argued against these overcompensation accounts in two ways. First, the pseudoword effect was demonstrated on a forced-choice recognition test. It is often assumed that forced-choice tests do not require response criteria, as participants simply select the stronger of the two alternatives. Participants were given a mixed list of words and pseudowords and then were given a forced-choice test in which all test pairs contained one word and one pseudoword. Although across test pairs the word and the pseudoword were equally likely to be the correct recognition response, participants chose the pseudoword 62% of the time. Second, the pseudoword effect was found in situations where it was obvious to participants that pseudowords were not less memorable than words and that therefore no rescaling or criterion-movement would be needed; this was done either by enhancing encoding of pseudowords by repetition or by degrading encoding of words using directed-forgetting instructions for words alone. Participants were asked about their expectations prior to the recognition test, and all expressed the belief that pseudowords in these conditions would be better remembered than words. However, when tested for recognition, they continued to give more positive responses to pseudowords than to words, indicating that conscious beliefs about stimulus memorability were unlikely to play any role in these response differences.

Greene (2004) argued that the pseudoword effect reflected the fact that pseudowords tend to be more familiar than words. Familiarity is often assumed to be determined by the similarity of the test item to the study list as a whole determined through a process of global matching (see Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982 for early presentations of this concept, and Ratcliff & McKoon, 2000 for an overview). Words contain semantic features that help to distinguish them from each other. Semantic features reduce the similarity of words to each other, thereby, reducing their overall familiarity levels. In contrast, pseudowords are much more impoverished semantically. Although some may remind people of words and thereby gain some semantic content, pseudowords will clearly on the average have much less meaningful content than will words. The similarity of pseudowords to the study list would be mostly based upon structural features, such as letters and phonemes, which are likely to be shared with a large number of studied items. Because structural features overlap greatly between items, structurally based familiarity could be greater than meaning-based familiarity.

Many models of recognition memory (e.g., McClelland & Chappell, 1998; Hintzman, 1988; Murdock, 1997; Shiffrin & Steyvers, 1997) assume that recognition tests are actually disguised tests of similarity: that is, how experiment participants make a recognition decision is by calculating the similarity of the test items to the study list and then comparing this global similarity value (or a transformed version of it) to some decision criterion. Indeed, a pseudoword effect is found when participants study a list and are then merely asked to rate new words and pseudowords for similarity to the studied list (Greene, 2004); new pseudowords are rated as more similar to a studied mixed list of words and pseudowords than are new words. Critical to the pseudoword effect is that these stimuli are indeed physically similar to words (and, in particular, to the words occurring on the study list), allowing a sense of familiarity to result from this similarity. When pseudowords are constructed specifically so that they do not resemble common words, the pseudoword effect does not occur (Groninger, 1976; Whittlesea & Williams, 2000). When words are intermixed on a list with nonwords that do not follow standard orthographic rules (e.g., the nonwords are random strings of consonants), nonwords do not receive more positive responses than words; a mirror pattern is found instead (Greene, 1996; Greene & Thapar, 1994). Direct evidence for the importance of structural similarity in determining familiarity was provided by Ozubko and Joordens (2008). Because words and pseudowords place constraints on the extent to which similarity can be manipulated, these investigators used cartoon characters as experimental stimuli and showed that manipulations of similarity can move the relationship of hits and false alarms from a mirror pattern to a concordant pattern. Increased similarity led to higher hit and false-alarm rates.

The greater number of positive responses given to pseudowords typically coexists with an overall memory advantage for words, at least in the absence of differential encoding. Ozubko and Yonelinas (2012) showed that this was not necessary by selecting words and pseudowords from much larger sets so that the stimuli used could be equated for overall performance. When words and pseudowords are chosen so that they are equally memorable, a recollection advantage is found for words and a familiarity advantage is found for pseudowords. However, participants continued to give more positive responses to pseudowords than to words.

I am arguing here that pseudoword effects reflect the enhanced similarity of pseudowords to a studied list containing words and pseudowords. This enhanced similarity leads to a greater sense of familiarity that occurs as a result of, rather than in spite of, the impoverished semantic features of pseudowords. Words have unique semantic elements that can distinguish between stimuli that may be quite similar in orthography (e.g., *chain* and *chair*). This reduces the average interitem similarity of words. Pseudowords composed from a common set of letters and phonological features and limited in semantic content will be high in interitem similarity, leading to enhanced familiarity. Ozubko and Joordens (2011) tried to simulate the pseudoword effect with the Retrieving Effectively from Memory (REM) model of Shiffrin and Steyvers (1997) using several hypothetical mechanisms. The only successful simulation of the effect was one assuming that there were familiarity differences between words and pseudowords resulting from the lack of distinctive semantic features in pseudowords. In fact, this simulation suggested a rather counterintuitive prediction, namely, that extremely high-frequency words (e.g., *and, that, this, from,* and *them*) are also deficient in semantic features. A subsequent experiment demonstrated that extremely high-frequency words lead to recognition-memory results resembling those found with pseudowords, namely, elevated hits and false alarms accompanied by poor accuracy.

If the lack of semantic features can enhance familiarity in recognition, is it possible that an inordinately high number of semantic features can reduce familiarity? This has not been directly tested, but this possibility offers one explanation for an otherwise surprising pattern of results reported by Lindsay, Kantner, and Fallow (2015). These investigators tested recognition memory for reproductions of paintings and found that participants made far fewer hits and false alarms to these stimuli than to words. It is as if the sheer amount of meaningful detail in a painting differentiates each one from all of the others on the study list. This could have the effect of reducing the similarity of each painting probe to the study list as a whole and thereby lowering familiarity. Too much meaning can be a bad thing in memory.

The pseudoword effect is a demonstration of how the removal of meaning can lead to an enhancement of the memory experience. That is, reading a pseudoword, a stimulus with typical orthographic structure but impoverished semantic features, subjectively makes you feel like you have seen it recently, whether or not you have encountered it before. Of course, the pseudoword effect also illustrates how important semantic features are, as genuine words are typically recognized far more accurately than are pseudowords. On a typical word list, semantic features may be used to differentiate between stimuli that may share many perceptual features, thereby, leading some words to be perceived as clearly old and others to be perceived as clearly new. In the absence of semantic features, the physical structure of the items may drive the familiarity process so that positive recognition responses become more likely whether or not they are appropriate.

3. WHAT WE KNOW AND WHAT WE DON'T KNOW

3.1 Empirical Overview

Long-term retention is often presented as a primarily semantic process, driven largely by the amount of meaningful processing that had been carried out on stimuli. Structural aspects of stimuli are commonly seen as having only the most transient effects on memory performance. This view of retention is laid out most explicitly in the levels-of-processing framework (Craik & Lockhart, 1972), although similar assumptions had been made by both earlier dual-store theorists and later processing theorists. Three lines of research have been reviewed that challenge this emphasis on the semantic base of retention. Orthographic distinctiveness, whether determined subjectively by ratings of "weirdness" or objectively through measures of orthographic neighborhood size, can influence the sort of processing that people carry out on words and therefore performance on a number of memory measures. Participants are able to use word fragments as effective cues for memory retrieval even when retrieval of word identity fails. When semantic features are impoverished and participants have to rely on phonological or orthographic features alone when making recognition judgments, this can paradoxically boost the experience of familiarity and therefore the probability of making a positive recognition response.

3.2 Known Unknowns

All of the phenomena discussed in this chapter are classic verbal-learning list-learning experiments. Many psychologists hope to draw lessons for educational practice from such experiments (see Marsh & Butler, 2013 for a review of such attempts). Still, work on physical structure and memory would appear to be particularly unpromising as a source for instructional practice. Although some education involves the mastery of basic facts, most instruction aims to teach students to draw and retain meaningful inferences. It is difficult to see how any of the work here can easily be applied to educational settings. General principles of memory, such as the emphasis on distinctiveness and retrieval cueing or the distinction between recollection and familiarity, may be illustrated through the experiments reviewed here but occur repeatedly throughout the memory literature.

It is not even clear how widely we can generalize from these findings to other verbal-learning experiments. All of the experiments reviewed here use carefully chosen sets of stimuli. For example, let us consider the effects of orthographic distinctiveness on memory. Experimenters wishing to demonstrate such effects have been careful to create lists composed of equal numbers of two distinct sets, containing words that are either very high, or very low, in orthographic distinctiveness. We follow this strategy to maximize our likelihood of finding an effect. What is still unknown is whether orthographic distinctiveness effects occur in lists that are not carefully chosen to yield such effects. Having many structurally weird words might lead participants to focus on their orthography. The fact that these distinctiveness effects are more likely to occur in mixed-list designs than in pure-list designs (Hunt & Elliott, 1980) is certainly consistent with the idea that participants must notice the systematic variation among stimuli before being influenced by structure. On the other hand, orthographic distinctiveness is correlated with other variables, such as word frequency and word length, and it has been speculated that orthography may actually contribute to the effects of these heavily studied variables (Jalbert et al., 2011; Malmberg et al., 2002). If so, then orthographic distinctiveness may play a large role in retention. Indeed, correlational megastudies on recognition of monosyllabic (Cortese, Khanna, & Hacker, 2010) and disyllabic words (Cortese, McCarty, & Schock, 2015) suggest that orthographic distinctiveness makes an independent contribution to performance. At this point, it is possible to make plausible arguments that the literature overstates or understates the magnitude of orthographic effects. We simply do not know the extent to which orthographic distinctiveness influences encoding on lists of random words in other memory experiments.

Recognition without identification shows that participants are able to use orthographic cues to access episodic information even when retrieval of the complete word fails. This is a striking effect, but it sheds no light on whether participants typically use orthographic information in recognition in experiments where fragments are not presented on the test. After all, in typical recognition experiments, there is a wealth of information that can be used to access episodic information, including semantic features. Moreover, most of the experiments demonstrating recognition without identification use carefully chosen stimuli (e.g., words that have unique two-word fragments) that are likely to have unusual structural properties. Whether orthographic cues are of much use in recognition of words that do not have such properties and are less physically distinctive is unclear.

The pseudoword effect shows that structural features can drive feelings of familiarity when semantic features are impoverished. Still, pseudowords in these experiments are not randomly created pronounceable nonwords but rather stimuli chosen specifically because they resemble real words. This may be an important element in these experiments, as Ozubko and Joordens (2008) found that interitem similarity is a critical variable here. The generality of the pseudoword effect beyond carefully chosen sets of stimuli has not been established.

The possibility that experiments on memory may not tell us much about memory outside the laboratory is well known (Neisser, 1982). However, we should keep in mind that psychological experiments with unusual procedures and carefully chosen stimuli may not even tell us much about other psychology experiments. Still, even though caution in generalization is warranted, what these experiments can reveal to us is the range of memory experiences we have. That is, although one does not have to believe that orthographic structure is critical for memory in all circumstances, the experiments discussed here demonstrate that it can have powerful effects on remembering in some circumstances. Memory is far more flexible than would be envisioned by approaches assuming that meaningful features always drive the processes of remembering.

3.3 Implications for Cognition

Traditionally, memory theories have embraced some form of abstraction. In the precognitive era, verbal-learning theories were often expressed in stimulus–response terms. Concepts such as reinforcement and habit strength were frequently employed, and it was commonly assumed that the responses in these experiments were determined by some process that blurs across multiple learning experiences (McGeoch, 1942). In the cognitive era, the abstraction process has been more explicit, with a clear commitment to the persistence of meaning and the rapid loss of physical detail. This commitment extends beyond memory to the claim that "it is the independence of thought from perception and action that makes human cognition special" (Mahon, 2015, p. 172).

Research on the influence of structural features in memory is more consistent with an alternative vision, where experiences are remembered not as generalized meanings but as individual episodes rich in perceptual and physical experience and where repeated events are represented not as blended holograms but as separate entities rich in the details of their separate occurrences (Greene, 2008). Indeed, this research on the effects of perceptual structure brings contemporary research on memory closer to an embodiment perspective (Barsalou, 1999; Glenberg, 2015), in which cognition is formatted in terms of sensorimotor experience.

In recent years, several theorists have encouraged us to carry out a broader examination of the purpose of memory (e.g., Gallo & Wheeler, 2013; Glenberg, 2015; Nairne, 2015). The classroom and the psychological laboratory are about equally unrepresentative of the range of situations in which we experience remembering. Although there are times when we have to recall a list of items or regurgitate the meaning of a textbook passage, throughout our lives it is far more common for us to remember specific events we experienced. These experiences are rich in perceptual and structural detail; indeed, it may be the specificity of these details that we learn to use to tell one event from another. We may use the presence of perceptual and structural details to determine the source of our memories so that we can distinguish between reality and imagination (Johnson & Raye, 1981). Hopefully, an acknowledgment that structural detail may drive recognition in even the most artificial of laboratory settings may lead us to a deeper appreciation of how it contributes to remembering as we go through our lives.

3.4 Memory and the Search for Universal Laws

The claim that short-term retention reflects sensory and structural features and long-term retention is driven by semantic factors was implicitly accepted for years before it was articulated in explicit form by Craik and Lockhart (1972). Indeed, the levels-of-processing framework constitutes a striking exception to the long-term trend in which psychologists increasingly refrain from expressing universal laws (Beach, 2014; Teigen, 2002). One reason perhaps why we shy away from formulating general principles is that those who do so may get immediately criticized by their peers. Craik and Lockhart did not escape such criticism, as many commentators immediately questioned not only the details of their proposal but the desirability of presenting a general principle at all. Mandler (1979) argued that memory theorists should "abandon the promised land of simple principles and return to the complexities of the human mind" (p. 305). Baddeley (1978) stated the belief that "the most fruitful way to extend our understanding of human memory is not to search for broader generalizations and 'principles,' but is rather to develop ways of separating out and analyzing more deeply the complex underlying processes" (p. 150). Indeed, Tulving (1985) questioned whether any general laws were even possible as "no profound generalizations can be made about memory as a whole" (p. 385). In fact, Tulving (2007) later wondered if there were at least 256 distinct types of memory, a question that was asked partly in jest but that was also seriously intended to point out the impossibility of forming universal generalizations across all forms of memory.

The results of any memory experiment are driven by the interaction of variables related to the participants, the information, encoding, and retrieval, making any search for general principles (Roediger, 2008) a quest doomed to failure. Still, the greatest contribution that any theoretical statement can ever make to a field may be to inspire and provoke research, even if the eventual result is to demonstrate the inadequacy of the original statement.

We now know that long-term memory is inextricably linked to the structural detail of experience, but this conclusion is largely a result of the prod to researchers provided by claims to the contrary.

REFERENCES

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–105). New York: Academic Press.

Baddeley, A. D. (1978). The trouble with levels: a reexamination of Craik and Lockhart's framework for memory research. *Psychological Review, 85*, 139–152.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.

Beach, L. R. (2014). Reconsidering a science of psychology built on laws. *American Journal of Psychology, 127*, 1–18.

Challis, B. H., & Brodbeck, D. R. (1992). Levels of processing affects priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 595–607.

Cleary, A. M. (2014). The sense of recognition during retrieval failure: implications for the nature of memory traces. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 60, pp. 77–112). San Diego, CA: Elsevier Academic Press.

Cleary, A. M., & Greene, R. L. (2000). Recognition without identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1063–1069.

Cleary, A. M., & Greene, R. L. (2001). Memory for unidentified items: evidence for the use of letter information in familiarity processes. *Memory and Cognition, 29*, 540–545.

Cleary, A. M., & Greene, R. L. (2002). Paradoxical effects of presentation modality on false memory. *Memory, 10*, 55–61.

Cleary, A. M., & Greene, R. L. (2004). True and false memory in the absence of perceptual identification. *Memory, 12*, 231–236.

Cleary, A. M., & Greene, R. L. (2005). Recognition without perceptual identification: a measure of familiarity? *Quarterly Journal of Experimental Psychology, 58A*, 1143–1152.

Cohen, R. L. (1985). On the generality of the laws of memory. In L.-G. Nilsson, & T. Archer (Eds.), *Perspectives on learning and memory* (pp. 247–277). Hillsdale, NJ: Erlbaum.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75–84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology, 55*, 429–432.

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science, 10*, 135–136.

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18*, 595–609.

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology, 68*, 1489–1501.

Cortese, M. J., Watson, J., Khanna, M., & McCallion, M. (2006). Revisiting distinctive processes in memory. *Psychonomic Bulletin and Review, 13*, 446–451.

Cortese, M. J., Watson, M. J., Wang, J., & Fugett, A. (2004). Relating distinctive orthographic and phonological processes to episodic memory performance. *Memory and Cognition, 32*, 632–639.
Craik, F. I. M., & Levy, B. A. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology, 86*, 77–82.
Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.
Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.
Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.
Crowder, R. G. (1970). The role of one's own voice in immediate memory. *Cognitive Psychology, 1*, 157–178.
Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica, 50*, 291–323.
Crowder, R. G., & Greene, R. L. (2000). Serial learning: cognition and behavior. In F. I. M. Craik, & E. Tulving (Eds.), *Handbook of memory* (pp. 125–135). Oxford, England: Oxford University Press.
Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics, 5*, 365–373.
Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology*. New York: Dover.
Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 671–682.
Gallo, D. A., & Wheeler, M. E. (2013). Episodic memory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 189–205). New York: Oxford University Press.
Gardiner, J. M., & Gregg, V. H. (1979). When auditory memory is not overwritten. *Journal of Verbal Learning and Verbal Behavior, 18*, 705–719.
Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.
Glanc, G. A., & Greene, R. L. (2007). Orthographic neighborhood size effects in recognition memory. *Memory and Cognition, 35*, 365–371.
Glanc, G. A., & Greene, R. L. (2009). Orthographic neighborhood size effects and associative recognition. *American Journal of Psychology, 122*, 56–61.
Glanc, G. A., & Greene, R. L. (2012). Orthographic distinctiveness and memory for order. *Memory, 20*, 865–871.
Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory and Cognition, 13*, 8–20.
Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100*, 546–567.
Glanzer, M., & Koppenaal, L. (1977). The effects of encoding tasks on free recall: stages and levels. *Journal of Verbal Learning and Verbal Behavior, 16*, 21–28.
Glanzer, M., & Schwartz, A. (1971). Mnemonic structure in free recall: differential effects on STS and LTS. *Journal of Verbal Learning and Verbal Behavior, 10*, 194–198.
Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 16–31.
Glenberg, A. M. (2015). Few believe the world is flat: how embodiment is changing the scientific understanding of cognition. *Canadian Journal of Experimental Psychology, 69*, 165–171.
Greene, R. L. (1985). Constraints on the long-term modality effect. *Journal of Memory and Language, 24*, 526–541.

Greene, R. L. (1996). Mirror effect in order and associative information: role of response strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 687–695.
Greene, R. L. (2004). Recognition memory for pseudowords. *Journal of Memory and Language, 50*, 259–267.
Greene, R. L. (2007). Foxes, hedgehogs, and mirror effects: the role of general principles in memory research. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 53–66). New York: Psychology Press.
Greene, R. L. (2008). Repetition and spacing effects. In H. L. Roediger III (Ed.), *Cognitive psychology of memory*. In J. Byrne (Ed.), *Learning and memory: A comprehensive reference* (Vol. 2, pp. 65–78). Oxford: Elsevier.
Greene, R. L., & Pearlman, I. (1996). The effects of vocalization on situational frequency estimation. *Memory, 4*, 453–460.
Greene, R. L., & Thapar, A. (1994). Mirror effect in frequency discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 946–952.
Greene, R. L., Thapar, A., & Westerman, D. L. (1998). Effects of generation on memory for order. *Journal of Memory and Language, 38*, 255–264.
Groninger, L. D. (1976). Predicting recognition during storage: the capacity of the memory system to evaluate itself. *Bulletin of the Psychonomic Society, 7*, 425–428.
Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551.
Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality tags and memory for repetitions: locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior, 12*, 229–238.
Hintzman, D. L., & Curran, T. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General, 126*, 228–247.
Hockley, W. E., & Niewiadomski, M. W. (2001). Interrupting recognition memory: tests of a criterion-change account of the revelation effect. *Memory and Cognition, 29*, 1176–1184.
Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior, 11*, 534–537.
Hunt, R. R., & Elliott, J. M. (1980). The role of nonsemantic information in memory: orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General, 109*, 49–74.
Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.
Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does word length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 338–353.
Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67–85.
Kang, S. H. K., Balota, D. A., & Yap, M. J. (2009). Pathway control in visual word processing: converging evidence from recognition memory. *Psychonomic Bulletin and Review, 16*, 692–698.
Lindsay, D. S., Kantner, J., & Fallow, K. M. (2015). Recognition memory response bias is conservative for paintings and we don't know why. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger, III (Eds.), *Remembering: Attributions, processes, and control in human memory, essays in honor of Larry Jacoby* (pp. 213–229). New York: Psychology Press.
MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 671–685.
Mahon, B. Z. (2015). The burden of embodied cognition. *Canadian Journal of Experimental Psychology, 69*, 172–178.

Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory and Cognition, 30*, 607–613.
Mandler, G. (1979). Organization and repetition: organizational principles with special reference to rote learning. In L. G. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary* (pp. 293–327). Hillsdale, NJ: Erlbaum.
Marsh, E. J., & Butler, A. C. (2013). Memory in educational settings. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 299–317). New York: Oxford University Press.
McCabe, D. P., Geraci, L., Boman, J. K., Sensenig, A. E., & Rhodes, M. G. (2011). On the validity of remember-know judgments: evidence from think aloud protocols. *Consciousness and Cognition, 20*, 1625–1633.
McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: a subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*, 724–760.
McDaniel, M. A., Cahill, M., Bugg, J. M., & Meadow, N. G. (2011). Dissociative effects of orthographic distinctiveness in pure and mixed lists: an item-order account. *Memory and Cognition, 39*, 1162–1173.
McGeoch, J. A. (1942). *The psychology of human learning*. New York: Longmans, Green.
Mickes, L., Searle-Carlisle, T. M., & Wixted, J. T. (2013). Rethinking familiarity: remember/know judgments in free recall. *Journal of Memory and Language, 68*, 333–349.
Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*, 609–626.
Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104*, 839–862.
Nairne, J. S. (2015). Adaptive memory: novel findings acquired through forward engineering. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger, III (Eds.), *Remembering: Attributions, processes, and control in human memory, essays in honor of Larry Jacoby* (pp. 3–14). New York: Psychology Press.
Nairne, J. S., Riegler, G. J., & Serra, M. (1991). Dissociative effects of generation on item and order information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 702–709.
Neisser, U. (1982). *Memory observed: Remembering in natural contexts*. San Francisco: Freeman.
Ozubko, J. D., & Joordens, S. (2008). Super Memory Bros: going from mirror patterns to concordant patterns via similarity enhancement. *Memory and Cognition, 36*, 1391–1402.
Ozubko, J. D., & Joordens, S. (2011). The similarities (and familiarities) of pseudowords and extremely high-frequency words: examining a familiarity-based explanation of the pseudoword effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 123–139.
Ozubko, J. D., & Yonelinas, A. P. (2012). A familiar finding: pseudowords are more familiar but no less recollectable than words. *Journal of Memory and Language, 66*, 361–375.
Peynircioglu, Z. F. (1990). A feeling-of-recognition without identification. *Journal of Memory and Language, 29*, 493–500.
Ratcliff, R., & McKoon, G. (2000). Memory models. In E. Tulving, & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 571–581). New York: Oxford University Press.
Raymond, B. (1969). Short-term storage and long-term storage in free recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 567–574.
Roediger, H. L., III (1990). Implicit memory: retention without remembering. *American Psychologist, 45*, 1043–1056.
Roediger, H. L., III (1993). Learning and memory: progress and challenges. In D. E. Meyer, & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 509–528). Cambridge, MA: MIT Press.
Roediger, H. L., III (2008). Relativity of remembering: why the laws of memory vanished. *Annual Review of Psychology, 59*, 225–254.

Roediger, H. L., III, Stadler, M. L., Weldon, M. S., & Riegler, G. L. (1992). Direct comparison of two implicit memory tests: word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1251–1269.
Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: a processing account. In H. L. Roediger, III, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.
Ryals, A. J., & Cleary, A. M. (2012). The recognition without cued recall phenomenon: support for a feature-matching theory over a partial recollection account. *Journal of Memory and Language, 66*, 747–762.
Seamon, J. G., & Murray, P. (1976). Depth of processing in recall and recognition memory: differential effects of stimulus meaningfulness and serial position. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 680–687.
Selmeczy, D., & Dobbins, I. G. (2014). Relating the content and confidence of recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 66–85.
Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM – retrieving effectively from memory. *Psychonomic Bulletin and Review, 4*, 145–166.
Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin, 75*, 399–415.
Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin and Review, 5*, 710–715.
Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1379–1396.
Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior, 1*, 443–450.
Teigen, K. H. (2002). One hundred years of laws in psychology. *American Journal of Psychology, 115*, 103–118.
Thapar, A., & Greene, R. L. (1994). Effects of levels of processing on explicit and implicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 671–679.
Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. Lewiston, NY: Macmillan.
Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*, 385–398.
Tulving, E. (2007). Are there 256 different kinds of memory? In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 39–52). New York: Psychology Press.
Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72*, 89–104.
Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: the discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 547–565.
Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: 1. The heuristic basis for feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 3–13.
Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 681–690.
Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin and Review, 11*, 616–641.
Yonelinas, A. P., & Jacoby, L. L. (2012). The process-dissociation approach two decades later: convergence, boundary conditions, and new directions. *Memory and Cognition, 40*, 663–680.

This page intentionally left blank

CHAPTER FIVE

The Role of Motor Action in Memory for Objects and Words

René Zeelenberg[1] and Diane Pecher
Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands
[1]Corresponding author: E-mail: zeelenberg@fsw.eur.nl

Contents

Abstract

Behavioral and neuroimaging data suggest that the actions associated with objects and words are automatically activated during object and word recognition. For example, recognition of a hammer activates the grip that is used to grasp a hammer and the actions that are performed when using a hammer. The question addressed in this review is whether these motor simulations support short-term and long-term memory for objects and verbs that are associated with actions. A meta-analysis shows that there is no evidence supporting a role for motor simulations in short-term recognition and *n*-back tasks. Serial recall tasks, on the other hand, have provided evidence that motor

Psychology of Learning and Motivation, Volume 64
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2015.09.005

simulations support short-term memory. The majority of these studies, however, used procedures that emphasized actions. These studies do therefore not provide strong evidence for the view that motor actions are automatically activated and encoded in memory. More studies are needed to establish a role for motor actions in short-term memory when actions are not primed by the context of the experiment. Only a few studies have studied the role of motor simulations in long-term memory. The available evidence suggests that motor simulations do not affect encoding in long-term memory. Overall the results cast some doubt on the idea that action has a central role in cognitive processing.

1. INTRODUCTION

When people perceive a stimulus, they activate much more information than is actually present in the stimulus itself. For example, when a printed word is presented a phonological representation is activated even when there is no intention to say the word out loud. This phonological representation is activated even when the word is presented subliminally and some have argued that the phonological representation is not a mere by-product of visual word processing, but rather plays an important role in meaning access (van Orden, 1987). Semantic priming studies suggest that the meaning of a word is also automatically activated (e.g., den Heyer, Briand, & Dannenbring, 1983; Neely, 1977; Pecher, Zeelenberg, & Raaijmakers, 2002). Some recent studies even suggest that rather specific information related to the use of objects may be automatically activated when people view (pictures of) objects or even words referring to objects (e.g., Tucker & Ellis, 1998, 2004). If object perception and word reading automatically activate associated actions, the question arises whether this information is stored in memory and supports later memory for objects and words. It seems reasonable to assume that memory would benefit from storing the actions that are associated with objects. Including actions in memory would result in a richer memory trace, increasing the number of cues that could potentially be used to retrieve items and making individual items more distinguishable from each other. Moreover, if the actions that can be performed with an object are activated automatically, they might be encoded in memory. Finally, if our cognitive system has evolved to support our interactions with the environment, one might expect that actions play an important role in memory. In recent years, this question has started to gain attention in the literature. In this chapter, we provide an early review of these studies.

1.1 Grounded Cognition and Action

Initial theories of conceptual representations treated concepts as abstract, amodal entities. More recent theoretical views propose that conceptual representations are based on perceptual and motor experiences and share processing resources with perception and action (Barsalou, 1999; Glenberg, 1997). For example, the concept *coffee* consists of the color of coffee, its smell and taste, and also the mouthfeel and temperature associated with drinking coffee. On this view, a concept such as *coffee* is represented by running a perceptual simulation that involves the same perceptual systems that are involved in actual experiences with the concept. Evidence for this view was obtained, among others, by Pecher, Zeelenberg, and Barsalou (2003). They presented concept–property pairs in a property verification task and manipulated whether the properties on successive trials were from the same modality or not. Responses to concept–property pairs (e.g., *blender-loud*) were faster if the property on the previous trial was from the same modality (e.g., *leaves-rustling*) than if the property was from a different modality (e.g., *cranberry-tart*). Pecher et al. (2003) suggested that in perceptual simulations of concepts, analogous to switching costs observed in perception (Spence, Nicholls, & Driver, 2000), switching attention from one modality to another one incurs a switching cost. Many other studies have also obtained results consistent with the idea that perceptual simulations underlie conceptual representations (e.g., Marques, 2006; van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008; Zwaan, Stanfield, & Yaxley, 2002; see Pecher, 2013b for an overview).

An important aspect of the grounded cognition view is that cognition is for action (Glenberg, 1997; Glenberg, Witt, & Metcalfe, 2013). Glenberg proposed that concepts are the meshed representations of two things: the affordances that are perceived in a current situation and memories of previous actions in similar situations. The concept of a functional object such as a hammer thus is a combination of the affordances that are perceived in the current hammer (e.g., grasping the handle) and memories of previous actions that were performed with hammers. On this account, concepts support interactions with the environment by activating previous actions. This view suggests that other cognitive processes such as language understanding and memory, in the absence of the actual objects, also might rely on the activation of simulated actions.

Evidence in support of this view is that pictures of objects potentiate actions that are compatible with the objects (Tucker & Ellis, 1998, 2004).

For example, when participants have to categorize objects as natural or artifacts, responses are faster when the response grip (e.g., a pinch between thumb and index finger) matches the grip that would be used on the object (e.g., a paperclip) than when it does not match (e.g., a hammer). This effect is caused not only by the visible features of objects but also by knowledge of objects, because similar findings are obtained when the stimuli are pictures of objects or object names (Bub & Masson, 2010b; Bub, Masson, & Cree, 2008; Glover, Rosenbaum, Graham, & Dixon, 2004; Masson, Bub, & Lavelle, 2013; Masson, Bub, & Warren, 2008; Rueschemeyer, Lindemann, Rooij, van Dam, & Bekkering, 2010; Tucker & Ellis, 2004). Furthermore, Bub et al. (2008) showed that not only volumetric grasping responses were activated but also functional motor actions, and Masson, Bub, and Breuer (2011) showed that orientation congruency effects for grasping responses to object pictures depended on whether the resulting grip allowed functional use of the object. These two studies thus showed that actions that may not be directly perceived in the object are also activated, suggesting that motor actions are retrieved from memory. Neuroimaging studies have shown activation of brain areas that are associated with motor actions when participants process manipulable objects (Buccino, Sato, Cattaneo, Rodà, & Riggio, 2009; Chao & Martin, 2000; Creem-Regehr & Lee, 2005; Martin & Chao, 2001; Martin, Wiggs, Ungerleider, & Haxby, 1996; Rueschemeyer, van Rooij, Lindemann, Willems, & Bekkering, 2009). These findings support the idea that motor actions that are typically used during interactions with objects are also activated during mental representation of those objects (but see Proctor & Miles, 2014 for an alternative view), suggesting that concepts are supported by motor simulations.

During language processing, simulated motor actions seem to play a role in comprehension (Borghi & Riggio, 2009; Casteel, 2011; Myung, Blumstein, & Sedivy, 2006). For example, Klatzky, Pellegrino, McCloskey, and Doherty (1989) showed that priming a compatible hand configuration facilitated comprehension of sentences such as *squeeze a tomato*. Several other studies have shown that language comprehension activated response actions that were congruent with some aspect of the action described in the language, such as direction (Glenberg & Kaschak, 2002), rotation (Taylor, Lev-Avi, & Zwaan, 2008; Zwaan & Taylor, 2006), or the action goal (Lindemann, Stenneken, van Schie, & Bekkering, 2006). Holt and Beilock (2006) showed that sentence—picture matching was influenced by the participant's experience with the action described by the sentence, also suggesting that action simulation is based on memory for previous actions.

Activation of motor areas when participants read action verbs or sentences further suggests a role of motor knowledge for higher cognition (Hauk, Johnsrude, & Pulvermüller, 2004; Raposo, Moss, Stamatakis, & Tyler, 2009; Rueschemeyer, Brass, & Friederici, 2007; Saccuman et al. 2006; Tettamanti et al., 2005; Willems & Hagoort, 2007).

Although this evidence seems to overwhelmingly show that motor actions are fundamental for understanding concepts and language, most findings of motor activation do not necessarily reflect core features of the conceptual process itself. Rather, congruency effects might be the result of secondary activation (Bub et al., 2008; Mahon, 2015a, 2015b; Mahon & Caramazza, 2008; Masson, 2015; Page, 2006) that occurs after the concept has already been understood. In order to argue that motor actions are necessary for understanding, we need to show that when activation of motor actions is compromised, conceptual understanding suffers. In neuropsychology, however, data from patients with damage to the motor system do not always indicate major problems for conceptual processing (Mahon, 2015a). In behavioral studies, the evidence is mixed. Typically, participants are asked to perform a conceptual task on objects that are manipulable (e.g., tools) and on objects that are nonmanipulable (e.g., animals). For manipulable objects, motor actions should be important for understanding, while for nonmanipulable objects they should not. Following this logic, a concurrent motor task should interfere with the understanding of manipulable objects but not, or less, with the understanding of nonmanipulable objects. Some studies have indeed obtained evidence that motor interference is more disruptive for conceptual processing of manipulable than nonmanipulable objects. For example, when participants name or categorize object pictures, performance for manipulable objects is decreased when participants are using their grasping hand for an unrelated concurrent action, whereas performance for nonmanipulable objects does not suffer (Witt, Kemmerer, Linkenauger, & Culham, 2010; Yee, Chrysikou, Hoffman, & Thompson-Schill, 2013; see also Rueschemeyer, Lindemann, Rooij, Dam, & Bekkering, 2010). Other studies, however, have failed to find effects of motor interference on conceptual processing (Matheson, White, & McMullen, 2014a, 2014b; Postle, Ashton, McFarland, & de Zubicaray, 2013). In sum, although more research might be needed to establish to what extent motor interference disrupts conceptual processing, several studies have shown that concepts and motor actions are strongly related.

Given this importance of motor actions for concepts and language understanding, we asked what the role of motor actions would be for more explicit types of memory, such as short-term memory and long-term episodic memory. It has been long established that episodic memory is influenced by conceptual variables such as category membership. For example, the release from proactive interference effect shows that interference between items in memory depends on conceptual overlap between items (Marques, 2000; Wickens, 1970; Wickens, Dalezman, & Eggemeier, 1976; Zinober, Cermak, Cermak, & Dickerson, 1975) and memory for word lists is affected by the conceptual organization of the study list (Lewis, 1971). These, and many other findings (e.g., Barclay, Bransford, Franks, McCarrel, & Nitsch, 1974; Deese, 1959; Hemmer & Steyvers, 2009; Light & Carter-Sobell, 1970), show that people rely on conceptual information to store items in and retrieve items from memory. Assuming that motor actions form a part of conceptual knowledge, motor actions should play an important role in memory.

2. SHORT-TERM MEMORY

Below we will first describe recently published studies examining the role of motor affordances in short-term or working memory.[1] The number of published studies on this topic is still small, especially when compared to the large number of studies examining the online activation of affordances by objects and words. We take advantage of this by describing the studies in relative detail. We will then evaluate whether the results of these studies support a role for motor simulations in short-term memory for objects and words. We will also examine the factors, if any, that modulate when motor affordances do and do not play a role in maintaining representations in short-term memory.

2.1 Neuroimaging Evidence

Mecklinger, Gruenewald, Weiskopf, and Doeller (2004) investigated the role of motor affordances in visual working memory for manipulable and nonmanipulable objects in an functional magnetic resonance imaging (fMRI) study. On each trial a line drawing of an object was presented on the screen for 100 ms. After 4 s a task cue was presented, which indicated the task to be performed on the test stimulus which was presented 6 s after the task cue. The task consisted of either a memory task or a control task. In the memory task,

[1] We use the terms short-term memory and working memory interchangeably.

participants decided if the test stimulus was identical to the studied object or whether the test stimulus was a mirror image of the studied object. In the control task, participants decided whether two digits, presented to the left and right of the test object, were identical or not. Mecklinger et al. inspected activation patterns on memory trials relative to control trials. For manipulable objects, but not for nonmanipulable objects, enhanced activation of the left ventral premotor cortex (PMC) was observed. For nonmanipulable items, enhanced activation was found in the left inferior frontal gyrus (Broca's area). These findings suggest that motor programs play a role in the maintenance of manipulable objects in working memory, whereas speech programs play a role in the maintenance of nonmanipulable objects. In a subsequent experiment, participants performed two different memory tasks (and the control task). In the movement comparison task, participants decided whether the test stimulus (e.g., *fork*) and study stimulus (e.g., *spoon*) afforded similar hand movements. In the size comparison task, participants decided whether the test stimulus (e.g., *whistle*) and study stimulus (e.g., *lipstick*) were of similar size. The primary result of this experiment was that activation of the ventral PMC was enhanced in the movement comparison task relative to the size comparison task. This result suggests that the involvement of the motor system in maintaining objects in working memory is to some degree dependent on task requirements.

A number of factors complicate interpretation of these results. First, the data analyses were performed on only a subset of the stimuli (high symmetry objects, but not low symmetry objects). Second, each object was presented only twice in the experiment. In working memory experiments, a small set of stimuli is typically presented multiple times to limit contributions from long-term memory. When stimuli are presented only once or twice in the experiment, a significant part of performance may be based on retrieval from long-term memory. Third, Mecklinger et al. (2004) also obtained activation of the left ventral PMC in a passive viewing task. The activation of the left ventral PMC can therefore not be unambiguously attributed to the maintenance of objects in working memory. Finally, drawing causal conclusions about mental processes from fMRI results is problematic (e.g., Page, 2006; Poldrack, 2008). One problem is that the observation that a specific brain region is active during the performance of a certain task does not imply that the specific region is necessary or sufficient for performing the task.

Thus, the results of Mecklinger et al. (2004) are consistent with the idea that motor simulations play a role in maintaining object representations in working memory. Alternative interpretations, however, are that the results reflect contributions from long-term or semantic memory, instead of short-term memory, or that activation of the PMC was merely epiphenomenal. In the next section we discuss behavioral studies, inspired by the findings of Mecklinger et al., that attempted to establish a causal role for motor simulations in short-term memory.

2.2 Motor-Interference Effects

The results of Mecklinger et al. (2004) showed enhanced activation of motor areas when people had to maintain manipulable objects in short-term memory. This suggests the possibility that the motor system plays a functional role in short-term memory for objects. If indeed the motor system contributes to memory for objects, interfering with the motor system should negatively affect memory performance. Witt et al. (2010) (but see Matheson, White, & McMullen, 2014b) found that a motor-interference task affected naming latencies for pictures of tools, but not for pictures of animals. They argued that the concurrent motor task interfered with the ability to form a grasping simulation and that motor simulations play a functional role in tool identification. Moreover, Smyth and Pendleton (1989) showed that working memory span for sequences of hand configurations was decreased by a concurrent task that changed the hand configuration (squeezing a tube). Likewise, if motor simulations support working memory for words and objects, concurrent motor-interference tasks are expected to affect working memory performance.

2.2.1 Effects of Interfering Actions on Short-Term Recognition Memory for Objects

Several studies have investigated whether motor-interference tasks differentially affect performance in short-term memory tasks for stimuli that differ in the way they are associated to motor actions. Pecher (2013a) studied the role of motor affordances in a short-term recognition task. On each trial, a stimulus was presented and after a 5000-ms retention interval the test stimulus was presented. Participants had to decide whether the test stimulus was identical to the studied stimulus or not. During the retention interval, participants performed a motor-interference task, a verbal-interference task, both tasks simultaneously, or no task. In the motor-interference task, participants started by making a fist with both hands. They then opened their fist by stretching their fingers one by one (simultaneously for both hands), starting with their thumbs, until their hands were completely opened.

Subsequently, participants made two fists again and opened their fingers in the same prescribed pattern, and so on. Importantly, this interfering task resulted in continuous changes in the hand configuration which were expected to interfere with activation of motor actions related to grasping and using objects. In the verbal-interference task, participants repeated four nonsense syllables (*bah-doh-ree-su*) out loud. The stimuli were pictures of manipulable (e.g., *binocular, corkscrew, calculator*) and nonmanipulable objects (e.g., *parakeet, painting, chimney*). Because manipulable objects, but not nonmanipulable objects, are associated with motor actions the prediction was that a concurrent motor-interference task would be particularly detrimental to performance for manipulable objects. However, no interaction was found between type of object and type of interference task. To be clear, performing concurrent interference tasks did negatively affect performance relative to the no-task control condition, but not differentially for manipulable and nonmanipulable objects. Across five experiments, similar results were obtained with different types of distractors ("new" nonstudied stimuli or mirror images of the presented objects), different types of stimuli (pictures of objects or words referring to objects), and different memory loads (one or four stimuli were presented during study).

In a follow-up study, Pecher et al. (2013) presented manipulable and nonmanipulable objects in an *n*-back task (rather than a short-term recognition task). The lag varied from 1 to 4 and participants performed the same concurrent interference task as in Pecher (2013a). Performing concurrent interference tasks and longer lags negatively affected memory performance, but again there was no interaction between type of stimulus and type of interference task.

Quak, Pecher, and Zeelenberg (2014) manipulated the congruency of the studied objects and the concurrent motor task. Participants continuously performed a precision grip movement (squeezing and releasing a small foam rubber cylinder between the thumb and index finger), a power grip movement (squeezing and releasing a large foam rubber cylinder with the whole hand), or no concurrent task while performing a 3-back task. The stimuli consisted of pictures of nonmanipulable objects (e.g., *chimney, bridge*), manipulable objects that required a precision grip when interacted with (e.g., *needle, paper clip*), and manipulable objects that required a power grip when interacted with (e.g., *hammer, axe*). Again, it was expected that nonmanipulable objects would suffer less from performing a concurrent motor task than manipulable objects. Moreover, if the maintenance of object representations in short-term memory is supported by motor affordances, performing an incongruent motor task (e.g., performing a power grip movement while remembering objects that are associated with a precision grip) should result in worse performance than performing a congruent concurrent motor task. However, no interaction between type of object and concurrent task was found. Again, performing a concurrent motor task relative to performing no concurrent task had a detrimental effect on performance (but equally for all conditions, as indicated by the lack of an interaction). Similar results were obtained irrespective of whether the different types of object pictures were blocked (Experiment 1) or whether the presentation of manipulable and nonmanipulable objects was mixed (Experiment 2).

2.2.2 Effect of Limb Movements on Short-Term Memory for Action Verbs

In contrast to this series of studies, Shebani and Pulvermüller (2013) obtained evidence suggesting that the motor system is involved in short-term memory for words. Shebani and Pulvermüller presented arm-related action words (*grasp, clap*) and leg-related action words (*hike, kick*) for study in a serial order recall task. On each trial, four words (either all arm-related action words or all leg-related action words) were presented one at a time for 100 ms, followed by a 400-ms inter-stimulus interval. Stimulus presentation was followed by a 6-s retention interval after which the four words had to be recalled in the order in which they had been presented. Crucially, during the retention interval, participants performed one of four interference tasks: moving with the hands, moving with the feet, articulatory suppression, or no task. Of primary interest were the hand and foot movement conditions. In these conditions, participants tapped a drumming sequence alternating between the right and left known as the single paradiddle (RLRRLRLL) with their hands or their feet. Participants made more errors in the immediate serial recall task when the type of action word was concordant with the motor-interference task performed during the retention interval than when the type of action word and motor-interference task were not concordant. Thus, tapping a paradiddle with the hands was particularly detrimental for arm-related action words, whereas tapping a paradiddle with the feet was particularly detrimental for leg-related action words. These findings are consistent with the view that tapping the paradiddle with the hands interferes with forming a motor simulation for arm-related action words that supports the maintenance of a representation of the action word in memory. Likewise, tapping the paradiddle with the feet interferes with forming a motor simulation for leg-related action words. Thus, in contrast to the studies from our lab with object pictures, the results of Shebani and Pulvermüller suggest that motor simulations play a role in working memory for action words.

Note that there are many differences between the experiment by Shebani and Pulvermüller (2013) and those of Pecher (2013a), Pecher et al. (2013) and Quak et al. (2014), in particular in the motor-interference tasks, stimuli and memory tasks being used. To uncover the critical differences responsible for these different findings, Zeelenberg and Pecher (2015b) performed two experiments. In a first experiment we examined whether the difference in results might be due to Shebani and Pulvermüller's (2013) use of the single paradiddle as a motor-interference task. The paradiddle might interfere more with the activation and storage of motor-interference in memory for several reasons. First, more so than the motor-interference tasks used by Pecher (2013a), Pecher et al. (2013), and Quak et al. (2014), the single paradiddle requires participants to closely monitor performance in the motor-interference task. Second, adequately performing the single paradiddle requires some practice to learn the correct sequence of alternating movements of the two hands or feet. Third, the speed at which participants performed the motor-interference task during the main experiment was individually adjusted depending on performance during the practice phase so that each participant did the task at the maximum speed at which they could still correctly perform the paradiddle.

In Experiment 1 of Zeelenberg and Pecher (2015b), the same materials and memory task were used as in Pecher et al. (2013), but the interference task was the single paradiddle used by Shebani and Pulvermüller (2013). Participants performed an *n*-back task for manipulable and nonmanipulable pictures of objects. Simultaneously, they performed the single paradiddle with their hands or feet. If motor simulations play a role in working memory, tapping the paradiddle with the hands as compared to tapping the paradiddle with the feet should be particularly detrimental for manipulable objects because, in contrast to nonmanipulable objects, manipulable objects are associated with motor actions (with the hands). That is, an interaction between stimulus type and type of motor-interference task should be found. No such effect was found, however, suggesting that Shebani and Pulvermüller's success in finding differential effects of motor interference on memory performance was not simply due to the single paradiddle being more interfering than the motor tasks used in other studies.

In a second experiment, we used action verbs related to the arms and legs, just as Shebani and Pulvermüller (2013). As in the first experiment, an *n*-back task was used and participants performed the single paradiddle with their hands or their feet. The interaction between stimulus type (arm-related action verbs vs leg-related action verbs) and effector (hand vs foot paradiddle) was not significant. Thus, unlike Shebani and Pulvermüller, our experiment failed to provide evidence for a role of the motor system in short-term memory for actions words.

Both experiments of Zeelenberg and Pecher (2015b) were designed to bridge the procedural gap between earlier experiments performed in our lab and the procedure used by Shebani and Pulvermüller (2013). Our results indicate that Shebani and Pulvermüller's findings were not likely caused by their use of the single paradiddle as a motor-interference task or their use of action verbs rather than pictures of objects. A more plausible aspect responsible for the different findings is therefore that different memory tasks were used in the different studies. We have used tasks such as short-term recognition and *n*-back tasks in our lab, whereas Shebani and Pulvermüller used a serial order recall task. We are currently planning an exact replication of Shebani and Pulvermüller (2013) in which we use a serial order recall task. Assuming that we replicate their findings, this would suggest that recalling items in the order in which they were presented is a critical factor is obtaining action-based memory effects. The findings that we discuss next are consistent with this idea.

2.2.3 Effects of Concurrent Motor Actions on Memory for Order

Evidence for a role of the motor system in short-term memory was also obtained by Lagacé and Guérard (2015) who studied the effect of movement congruency on performance in an order reconstruction task. On each trial, six object pictures were presented and after presentation of the list, the six objects were simultaneously presented on the screen and participants had to indicate the order in which the objects had been presented. During study, each object picture was preceded by a 300-ms video displaying one of the three possible grips: (1) a *power grip* (the object is held against the palm of the hand and the fingers close toward the palm of the hand), (2) an *index—thumb grip* (a delicate grip requiring small force where the object is held between the index finger and the thumb), and (3) a *parallel extension grip* (the object is held between the thumb and the whole surface of the fingers, which are pressed tightly against each other).[2] Participants in the grasping condition had to perform the grip displayed in

[2] The descriptions of the grips were copied from Downing-Doucet and Guérard (2014). Note that the *index—thumb grip* is identical to what others refer to as *precision grip* (e.g., Pecher, 2013a; Tucker & Ellis, 2001).

the video during the presentation of the subsequent object picture. Thus during each trial, six videos were presented which were each followed by an object picture while the participant performed the grip presented in the video. In half of the trials the video displayed an action congruent with the object picture (e.g., a video displaying a power grip followed by a picture of a hammer). In the remaining trials the video displayed an action incongruent with the object picture (e.g., a video displaying an index—thumb grip followed by a picture of a hammer). Participants in the control condition watched the sequence of videos and object pictures, but did not perform actions during study. The results showed that, for participants performing grips during study, order memory was better for congruent trials than for incongruent trials. For participants in the control group, no congruency effect was found.

Similar findings were obtained in another experiment in which participants did not pantomime the grasping action but rather pantomimed the actions associated with the use of an object. On each trial a sequence of prime pictures (surrounded by a blue frame) and target (surrounded by a red frame) pictures was presented. In half of the trials the action associated with the prime picture was congruent with the action associated with the target picture (e.g., prime picture: *axe*, target picture: *hammer*). In the other half of the trials the action associated with the prime picture was incongruent with the action associated with the target picture (e.g., prime picture: *corkscrew*, target picture: *hammer*). Again serial order memory was better for congruent trials than for incongruent actions, but only for participants who pantomimed the actions associated with the primes, and not for participants in the control condition who simply watched the study sequence.

2.2.4 Discussion of Motor-Interference Effects

In sum, working memory studies that have used motor interference to study the role of the motor system for memory have shown mixed results. Although some studies have shown that working memory for manipulable objects or action verbs is decreased when participants perform interfering motor actions, others have not obtained any evidence for a role of the motor system. A potential explanation for this difference is that the serial recall task might be more sensitive to motor information than item recognition tasks, but at present this explanation has not been tested.

2.3 Similarity-Based Effects

The two studies that we discuss next were modeled after two well-known effects of similarity among items on short-term memory performance: the similarity effect and the isolation effect. Many studies have shown that serial order recall for lists of similar items is worse than for lists of dissimilar items. Baddeley (1966) (also see Conrad, 1964), for example, found that recall of the order in which words had been presented during study was much lower for acoustically similar words (e.g., *mad, man, mat,* ...) than for acoustically dissimilar control words (e.g., *cow, day, car,* ...). Likewise, Jalbert, Saint-Aubin, and Tremblay (2008) showed impaired order recall for similarly colored squares as compared to dissimilarly colored squares. The isolation effect refers to the finding that items that possess a feature or characteristic that sets it apart from other items on the list are better recalled than items that do not possess such a feature. For example, Cimbalo, Capria, Neider, and Wilkins (1977) showed that consonants that differed in size and color from other items on the study list were better remembered than nonisolated control items (i.e., items not differing in these characteristics from surrounding items). For similar findings in the recall of spatial information, see Guérard, Hughes, and Tremblay (2008).

2.3.1 Motor-Similarity Effect

Downing-Doucet and Guérard (2014) investigated the effect of motor similarity on performance in an order reconstruction task; the same memory task that was used by Lagacé and Guérard (2015). Participants studied pictures of objects that were each associated with two out of four possible actions: (1) a *power grip* (2) an *index—thumb grip*, (3) a *parallel extension grip*, and (4) a *fingers—thumb grip* (the object is in contact with most of the fingers and is held between the tip of the fingers and the thumb). For example, according to norms (Lagacé, Downing-Doucet, & Guérard, 2013) a small box is equally associated with a fingers—thumb grip and a parallel extension grip. Other objects were associated with different pairs of grips.

On each trial, Downing-Doucet and Guérard (2014) presented six different objects. Prior to each object, a 300-ms video was presented showing a hand making a grasping movement (see Lagacé & Guérard, 2015, Figure 1 for examples). On similar trials each video showed the same grasping movement; on dissimilar trials the videos showed different grasping movements.[3] An important aspect of the design was that the same objects were used in similar and dissimilar lists. What differed was the action "primed" by the video shown prior to each object picture. After

[3] Because four different grips were used, some grips were shown more than once in a trial but all four grips were present in each trial.

presentation of the study list, the six objects were simultaneously presented on the screen and participants had to indicate the order in which the objects had been presented. Downing-Doucet and Guérard (2014) found that performance in this order reconstruction task was better for dissimilar lists than for similar lists. The advantage for dissimilar lists was absent in a follow-up experiment in which participants performed a concurrent motor-interference task (continuously opening and closing the fist in the manner used by Pecher, 2013a) during study. This finding is consistent with the idea that the concurrent motor task interfered with the formation of motor simulations for the presented objects. The results of Downing-Doucet and Guérard (2014) and the previously discussed study by Lagacé and Guérard (2015) are consistent with the notion that motor affordances play a role in short-term memory for object pictures. We note, however, that in both studies videos of hands making a grasping movement were shown prior to each object picture. This aspect of the design may have primed the use of motor affordances and be (partially) responsible for their success in obtaining evidence consistent with a role of motor affordances in short-term memory (see later for more discussion on this issue).

2.3.2 Motor-Isolation Effect

In another study, by Guérard and Lagacé (2014), motor actions were not explicitly primed. Guérard and Lagacé studied motor-isolation effects in serial recall. They reasoned that, in a way similar to the isolation effects described earlier, motor features associated with the stimuli might modulate memory performance. They presented pictures of objects that are easy to pantomime (*saw, punching bag, trampoline*) and pictures of objects that are hard to pantomime (*bust, moon, honeybee*). Although similar to the manipulable versus nonmanipulable manipulation used by Pecher (2013a), this distinction is different. For example, a *trampoline* could be considered a low manipulable object because one does not usually lift a *trampoline*, move it around, or change its orientation like one might do with a *comb* or a *fork*. Nevertheless, a *trampoline* is strongly associated with actions such as jumping and somersaulting and was thus considered a high pantomime object.

Guérard and Lagacé (2014) presented high and low pantomime objects in homogeneous or heterogeneous lists. The homogeneous lists consisted of either seven high pantomime objects or seven low pantomime objects. In heterogeneous lists the fourth object in the list differed in pantomime level from the other objects in the list (e.g., the fourth object was a high pantomime object and all other objects in the list were low pantomime objects).

If a motor-isolation effect is present, the fourth object in a heterogeneous list should be better recalled than the same object in a homogeneous list. This was indeed found by Guérard and Lagacé (2014). Moreover, their Experiment 2 showed that the motor-isolation effect was not present when participants performed a concurrent motor-interference task (continuously opening and closing the fist) throughout the experiment (i.e., during both encoding and recall). In Experiment 3, participants studied pictures of animals and pictures of man-made constructions. A (semantic or visual) isolation effect was present even though participants performed a concurrent motor-interference task (as they did in Experiment 2). This finding was taken to indicate that a motor-interference task selectively eliminates isolation effects that are based on the actions associated with objects but not those that are based on other features. In our view, this study provides one of the more convincing cases for a role of the motor system in working memory. First, unlike some of the other studies, the isolation manipulation does not explicitly draw attention to motor actions. Therefore, the effect might show that motor actions are spontaneously activated when participants memorize objects. Second, the concurrent motor task eliminated the motor-isolation effect, suggesting that the isolation effect actually depended on the motor system.

2.4 Other Studies on the Role of Motor Affordances

Apel, Cangelosi, Ellis, Goslin, and Fischer (2012) investigated the role of affordances in what they call an instruction span task. Participants watched a 3 × 3 grid on a touch screen, whose cells were numbered from 1 (upper left cell) to 9 (lower right cell), with eight cups positioned at the margins of the grid (two cups each above, below, to the left, and to the right of the grid). The cups had uniquely colored handles and participants received auditory instructions indicating which cup had to be moved to what location (e.g., "Move the orange cup to square three, then move the green cup to square nine, then move the yellow cup to square eight"). The number of instruction components per trial (i.e., cups that needed to be moved) ranged from three to six. After the instructions had been delivered, participants dragged the cups from their position around the screen to the specified cell in the grid, by touching the cups on the screen with their index finger.

Apel et al. (2012) manipulated whether participants moved the cups with their right hand or their left hand (between subjects) and whether the cup handles in a trial were all oriented to the left, all oriented to the right or randomly mixed (within-subjects). In Experiment 1, all participants were right handed. The most important result was that participants who moved the cups on the screen with their right hand executed the to-be-remembered instructions more accurately when the cup handles were all oriented to the right than when they were all oriented to the left. This spatial congruency effect was expected because cups with handles oriented to the right would activate (pre)motor neurons that control movements with the right arm and hand. Such neurons, it was reasoned, might subsequently support recall and execution of the instructions. Unexpectedly, participants who moved the cups with their left hand showed no effect of handle orientation. In Experiment 2 all participants were left handed. This experiment showed no effect at all of handle orientation used during execution of the movements. Apel et al. (2012) attributed the absence of an effect for left-handed participants to the general design of man-made objects. Objects are usually designed with right-handed people in mind. As a result, left-handed people often use their nonpreferred hand when interacting with objects and therefore may not develop strong associations between objects and action representations. Although the action-based effect was not consistently obtained across conditions, these results provide some evidence for the notion that motor simulations play a role in immediate memory.

Pezzulo, Barca, Lamberti-Bocconi, and Borghi (2010) also obtained evidence consistent with the idea that affordances support immediate memory. Participants in their study were novice climbers (less than 6 months climbing experience) and expert climbers (between 5 and 10 years climbing experience). The climbers each studied three different climbing routes: a route that was relatively easy to climb, a route that was relatively hard to climb, and a route that was impossible to climb. Participants entered the climbing arena and the trainer twice indicated the route on the climbing wall. Participants then turned around, did a brief distractor task (saying the letters A to I) and where subsequently shown a paper sheet displaying a picture of the climbing wall. Participants marked the sequence of holds composing the route just studied. In accordance with the hypothesis of Pezzulo et al. expert climbers' recall performance was better than that of novice climbers for the difficult route. For the easy route and the impossible route recall performance of the two groups did not differ. These findings were predicted because both novice and expert climbers would be able to form a motor simulation for the easy route. For the impossible route, neither group would be able to form a motor simulation. For the difficult route, however, it was reasoned that only expert climbers would be able to form a motor simulation because expert climbers, but not novices, possess the necessary skills to perform the sequence of actions needed to climb the difficult route.

2.5 Evaluating Evidence for the Role of Motor Affordances in Short-Term Memory

Comparing studies that did and did not find evidence for a role of motor affordances in short-term memory, two factors seem to affect the presence of action-based memory effects. First, the studies that found no evidence for a role of the motor system used tasks in which participants were asked to decide whether or not a test stimulus had just been presented. Pecher (2013a) presented one or more stimuli on each trial and participants had to decide whether the test stimulus was identical to the study stimulus or not (i.e., a short-term recognition task). Pecher et al. (2013), Quak et al. (2014), and Zeelenberg and Pecher (2015b) all used *n*-back tasks. These tasks are variants of recognition tasks in which targets and distractors are presented for a memory judgment. In contrast, the studies that did find action-based memory effects (Apel et al., 2012; Downing-Doucet & Guérard, 2014; Guérard & Lagacé, 2014; Lagacé & Guérard, 2015; Pezzulo et al., 2010; Shebani & Pulvermüller, 2013) all used recall or recall-like tasks in which the studied stimuli themselves or the presentation order of the stimuli had to be retrieved from memory.

Before expanding on this distinction between memory tasks, we report the results of a meta-analysis that included all 11 experiments performed by Pecher (2013a), Pecher et al. (2013), Quak et al. (2014), and Zeelenberg and Pecher (2015b). Although all 11 experiments individually showed no evidence for a role of motor affordances in short-term memory, we wanted to find out if there might be evidence for such a role when the results of all experiments were combined into a single meta-analysis. For each experiment, we expressed the critical hypothesized interaction between stimulus type and type of interference task as a difference of the differences between conditions. Note that the interaction in a 2×2 repeated measures ANOVA is equivalent to a paired samples *t*-test on the difference of the differences. These difference scores and 95% confidence intervals for each experiment are shown in Figure 1.[4] For example, Experiment 1 of Pecher (2013a) showed that the motor-interference effect (i.e., the difference between conditions with and without concurrent motor-interference task) amounted to

[4] When calculating the difference scores that we entered in the meta-analysis, we collapsed over manipulations other than stimulus type and motor interference (i.e., verbal interference and lag manipulations, depending on the experiment). For the Quak et al. (2014) study the difference score was based on a comparison of the congruent and incongruent conditions only (excluding the results for nonmanipulable objects).

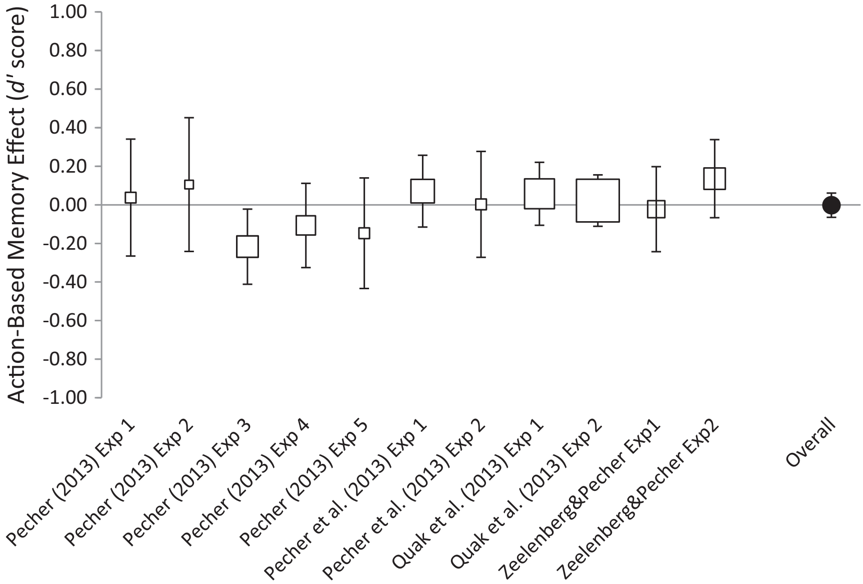

Figure 1 Results of the meta-analysis. The points show the size of the action-based memory effects for each experiment separately. The error bars show the 95% CI for each difference. The size of the markers reflects the weighing of each experiment in the meta-analysis. The overall point is the estimated effect size based on Cumming (2012).

0.394 d' points for manipulable objects and 0.356 d' points for nonmanipulable objects. Thus, the motor-inference effect was 0.037 d' points larger for manipulable objects than for nonmanipulable objects. This small and nonsignificant action-based memory effect is displayed as a positive effect in Figure 1 because it was in the direction predicted by the account that motor simulations support short-term memory for objects.

The meta-effect (indicated by the overall point) is a nonsignificant −0.002 difference in d' scores (95% confidence interval = [−0.065, 0.061]). The analysis showed no significant heterogeneity across experiments, $Q_{df=10} = 10.95$, $p = 0.361$, $I^2 = 8.69\%$. A perceptive reader may notice that Experiment 3 of the Pecher (2013a) study showed a significant effect in the direction opposite of what is predicted. Given that there was no significant heterogeneity across experiments, this most likely reflects random noise; as the number of experiments in a domain grows one is bound to occasionally find a significant effect, even when no true effect exists. Not surprisingly, an exploratory meta-analysis in which we excluded the results from Experiment 3 of the Pecher (2013a) study still showed a nonsignificant meta-effect

($M = 0.024$, 95% CI $= [-0.039, 0.086]$). The meta-analysis thus showed that in recognition-like short-term memory tasks, there is no evidence that motor simulations support memory performance.[5]

As we mentioned, the studies that did find action-based memory effects all used recall tasks. One question is whether it is the recall aspect itself that is responsible for the presence of an action-based memory effect or whether it is the requirement to recall order information. Note that almost all experiments required the recall of order information. Shebani and Pulvermüller (2013) and Guérard and Lagacé (2014) specifically required participants to recall the stimuli in the order in which they had been presented during study (and scored performance accordingly). Downing-Doucet and Guérard (2014) and Lagacé and Guérard (2015) presented the studied objects simultaneously on the computer screen during test and participants had to touch the objects in the order in which they had been presented during study. Pezzulo et al. (2010) required participants to recall the holds on the climbing hall in the correct sequence. Finally, Apel et al. (2012) scored the number of correctly executed action instructions regardless of the order in which they were performed. The instructions, however, implied that the actions had to be performed in a certain order (e.g., "Move the orange cup to square three, then move the green cup to square nine, then move the yellow cup to square eight"). It is conceivable that participants understood the instructions to indicate that the actions had to be executed in a particular order or at least tried to execute the actions in the order indicated by the instructions. Thus all these studies required serial order recall and most of them also required item recall during the test phase. One relevant question for future studies is therefore whether action-based memory effects in short-term memory are limited to tasks that test memory for serial order or whether such effects are also found in recall tasks in which serial order is irrelevant.

It is worth pointing out that effects of order recall and item recall can be dissociated. For example, Fallon, Groves, and Tehan (1999) found that in a serial order recall task phonological similarity negatively affected memory for the order in which the items were presented, but not memory for the items themselves. Lagacé and Guérard (2015) speculated that a serial order recall

[5] Ten out of eleven experiments showed a significant main effect of motor interference on memory performance. Thus motor interference negatively affects short-term memory, probably due to a central attentional bottleneck, but there is no interaction between stimulus type and motor-interference task.

task recruits the motor system more than the *n*-back and short-term recognition task used by Pecher (2013a), Pecher et al. (2013) and Quak et al. (2014) and argued that this is responsible for the absence of action-based memory effects in these tasks. This speculation is in line with ideas expressed by Acheson and MacDonald (2009) who argued that because actions are inherently sequential in nature (i.e., many actions cannot be performed in parallel), the motor system may be capable of supporting the serial ordering of responses (but see Engelkamp & Dehn, 2000). If it is indeed true that the motor system plays a role in the recall of order information, and if this is responsible for the findings of action-based memory effects in short-term memory, then no such effects should be found in recall tasks where order information is irrelevant.

A second potentially important factor for the presence of an action-based memory effect is whether or not the study or test procedures emphasized actions. Actions were emphasized in Downing-Doucet and Guérard (2014) and in Lagacé and Guérard (2015) by presenting videos of hand movements during the study phase prior to each to be remembered object picture. The testing procedure used in these studies in which the participants had to touch the objects on the screen to indicate the order in which they had been presented may also have contributed to the finding of an action-based memory effect. In the Apel et al. (2012) study, the memory task consisted of moving cups to the specified locations on the grid displayed on a touch screen. Finally, Pezzulo et al. (2010) asked climbers to recall routes on a climbing wall. Although it might be argued that the climbing routes could be retained purely on the basis of visual information, climbers routinely study climbing routes by mentally simulating climbing these routes. These simulations may include specific details such as the location and orientation of holds and the movements of arms, legs, and body. Moreover, the use of motor information in recall may have helped to constrain the possible sequence of holds. The finding of action-based memory effects in these studies is interesting, but they do not, in our view, answer the question of whether affordances routinely play a role in short-term memory for object pictures and words referring to objects or actions. If affordances are activated automatically and support retention of words and objects action-based memory effects should be obtained even when actions are not emphasized by the design of the study.

The claim that motor actions are activated automatically is a very strong one and may represent only very extreme views on the role of motor actions for cognition. Many studies have provided support for the notion that the context in which stimuli are presented has a substantial influence on which conceptual features are activated and encoded in memory (e.g., Barclay et al., 1974; Barsalou, 1981, 1993; McKoon & Ratcliff, 1995; Zeelenberg, 2005; Zeelenberg, Pecher, Shiffrin, & Raaijmakers, 2003). Pecher, Zeelenberg, and Barsalou (2004) investigated whether conceptual representations were affected by recent experiences with those concepts. Each concept (e.g., *apple*) was presented twice, with different properties, in a property verification task. Pecher et al. manipulated whether the properties presented with the concept on the two occasions were from the same modality or from different modalities. On the second presentation of a concept, participants responded faster and more accurately to the concept–property pair (e.g., *apple-green*) if the concept had been previously presented with a property from the same modality (e.g., *apple-shiny*) than if it had been previously presented with a property from a different modality (e.g., *apple-tart*). This finding was obtained even though multiple unrelated concept–property pairs intervened between the first and second presentation of a concept. According to Pecher et al. (2004), the simulation on the first presentation of the concept was focused on a specific modality. For example, verifying whether the property *green* is true for the concept *apple* results in a mental simulation that focuses on the visual modality. This simulation includes the relevant visual property that needs to be verified (*green*), but also other visual properties (e.g., *shiny, round*). On later trials, these visual properties are more readily available, resulting in a benefit for verifying same modality properties relative to different modality properties. Related findings have been found by Pecher, Zanolie, and Zeelenberg (2007) (also see Pecher, van Dantzig, Zwaan, & Zeelenberg, 2009; van Dantzig, Cowell, Zeelenberg, & Pecher, 2011). In a similar vein, simulations of motor actions might be context sensitive. It is difficult to specify what kinds of contexts would lead to spontaneous activation of motor actions that are still theoretically interesting. For example, in a series of experiments, Yu, Abrams, and Zacks (2014) failed to find alignment effects for pictures of objects with handles oriented to the left or the right. A spatial alignment effect was found only when participants were instructed to imagine picking up the object while making the upright-inverted decision. Yu et al. concluded that actions may be primed only to the extent that the action-relevant aspects of an object are emphasized. In this study, context was manipulated by giving an explicit instruction, and it would be hard to argue that a motor congruency effect in this case still shows evidence of such spontaneous activation. With more subtle manipulations, for example, requiring participants to use reach and grasp actions, the evidence might be more convincing (also see Bub & Masson, 2010a).

Although several of the published studies that have found evidence for action-based memory effects in short-term memory used procedures that emphasized motor actions and interactions with the to-be-remembered stimuli, this is not true of all experiments. Two exceptions are the experiment done by Shebani and Pulvermüller (2013) and the motor-isolation effect reported by Guérard and Lagacé (2014, Experiment 1). In these studies, no movies of videos of hands making a grasping movement were shown and participants did not have to touch or move objects on the computer screen. Nevertheless the results of these studies indicated that motor affordances support short-term memory for objects pictures and action words. In our view, the results of these two studies provide the strongest evidence to date for the view that motor simulations support short-term memory. Given that the number of studies providing strong support for a role of affordances in short-term memory is still limited, future studies will have to show if these findings can be replicated and extended to other stimuli and procedures.

3. LONG-TERM MEMORY

Compared to the large number of studies investigating the online activation of affordances, only a few studies have investigated the role of affordances in short-term memory. Even fewer studies have examined the role of the motor system in long-term memory for objects and words. Short-term memory is often assumed to rely largely on phonological or visuospatial representations (Baddeley, 2003). Semantic characteristics of stimuli seem to play a relatively small role in short-term memory (Baddeley, 1966). Long-term memory, on the other hand, is known to rely heavily on semantic representations (e.g., Barclay et al., 1974; Deese, 1959; Hemmer & Steyvers, 2009; Light & Carter-Sobell, 1970). One would therefore expect that motor simulations play a prominent role in long-term memory. Just as in working memory studies, long-term memory studies have investigated how performing motor actions affects memory for objects and words. Below we will first describe two studies that investigated the effect of motor actions performed during study, addressing the role of motor simulations in initial memory encoding. Subsequently we will describe a study that investigated the effect of motor actions performed after initial encoding, addressing the role of motor simulations in memory consolidation.

3.1 The Effect of Motor Actions on Memory Encoding

In a first study (Pecher, Wolters, Stolte, & Zeelenberg, 2015), we investigated free recall for pictures of manipulable and nonmanipulable objects. If action related information is automatically activated for manipulable objects, as suggested by many studies, it is reasonable to assume that this information is encoded into episodic memory traces and supports later memory for these objects. Interfering with the activation and encoding of action-related information by means of a motor-interference task is therefore expected to harm memory for manipulable objects. Participants in our study either performed the same motor-interference task that Pecher (2013a) used (i.e., repeatedly opening the fists by stretching their fingers one by one) or no task, in separate blocks. In each block, they studied a mixed list of manipulable and nonmanipulable object pictures. After a filler task, memory was tested in a free recall task in which participants named or described the studied pictures. In separate experiments, participants performed the motor-interference tasks during study, during recall, or during both study and recall. Performing a motor-interference task, as compared to the no task control condition, was expected to be particularly detrimental for memory for manipulable objects. However, in none of the experiments was there evidence for such an interaction. Thus, no evidence was obtained that motor simulations support long-term memory for objects.

A second study (Zeelenberg & Pecher, 2015a) was modeled after the Shebani and Pulvermüller (2013) study who showed that in serial order recall performing a motor-interference task with the hands versus the feet differentially affected error rates for arm-related and leg-related action words. In our experiment, participants studied mixed lists of arm-related and leg-related action words. In one block participants performed the single paradiddle with their hands, and in another block they performed the single paradiddle with their feet (the paradiddle order was counterbalanced). After each block a free recall task was given. Assuming that the effect of Shebani and Pulvermüller (2013) extends to long-term memory we should observe an interaction between word type and type of motor-interference task; that is, free recall should be worse for the concordant condition than the non-concordant condition. Contrary to this prediction, however, no such effect was obtained. Likewise, a second experiment did not show that free recall for manipulable objects was affected more by tapping the paradiddle with the hands (as compared to tapping the paradiddle with the feet) than free recall for nonmanipulable objects.

Thus, in a total of six experiments, we did not find any evidence that performing a concurrent manual motor task selectively affected free recall for manipulable objects or arm-related action words. These results suggest that motor actions are not spontaneously activated and encoded in memory when people study objects or words.

3.2 The Effect of Motor Actions on Memory Consolidation

In the only published study investigating action-based long-term memory effects, van Dam, Rueschemeyer, Bekkering, and Lindemann (2013) studied the effects of motor actions performed after initial learning of object names on subsequent memory performance. Participants first studied words under intentional learning instructions. The critical stimuli were words referring to objects that are associated with either a twisting action (*steering wheel, pepper mill, screw driver*) or a pressing action (*piano, remote control, doorbell*). After the study phase, participants performed a seemingly unrelated number-judgment task. Critically, during this intervening task participants responded to an irrelevant feature of stimuli that had not been presented during the study phase (i.e., whether a number of the screen was smaller or larger than 5). Responses were made either by means of a twisting action or a pressing action. The type of action performed was manipulated between subjects. Van Dam et al. reasoned that performing a motor action after initial learning would affect consolidation. More specifically, performing an action (e.g., responding by means of a twisting action) would enhance memory for congruent words (*steering wheel*) relative to incongruent words (*piano*).

Experiment 1 used a yes—no recognition task and the results showed enhanced recognition memory performance for congruent words relative to incongruent words. In Experiment 2, van Dam et al. (2013) presented words during the study phase and object pictures corresponding to the critical words during the test phase. The pictures were slowly demasked during presentation in the test phase. That is, picture presentation started with a completely black screen and 5% of the pixels became visible every 150 ms so that the picture gradually appeared out of the black mask. Participants had to indicate as quickly as possible when they identified the picture. Participants responded more accurately (but not faster) in the congruent condition than in the incongruent condition. In Experiment 3, a standard word fragment completion task was used during the test. Responses in the word fragment completion task were faster (but not more accurate) for congruent words than for incongruent words.

Although these results are interesting, the results of the implicit memory tasks are less convincing than they seem at first. For one thing, the analyses of Experiment 2 were based on performance averaged over both old (previously studied) and new (nonstudied) items. As van Dam et al. (2013) mention, one possibility is that the action performed during the intervening task biased retrieval in the subsequent picture-demasking task. Biased retrieval would operate on both studied and nonstudied items. Hence, results that include both studied and nonstudied items do not demonstrate that the motor actions performed after initial study affect consolidation processes that play a role in priming. It might just be that the retrieval for both old and new items was biased by the motor actions performed during the intervening task. To control for this possibility van Dam et al. reported an analysis that included only nonstudied items. These analyses showed no significant difference between congruent and incongruent items. This result is suggestive, but the appropriate analyses would have been to compare priming scores (i.e., the difference between the studied and nonstudied conditions) for congruent and incongruent items. Also, the word fragment completion task used in Experiment 3 is known to be susceptible to explicit retrieval processes (e.g., Reingold & Goshen-Gottstein, 1996), so it is not clear to what extend the results were based on implicit memory processes. It thus remains to be seen whether these implicit memory effects hold up under more tightly controlled conditions.

3.3 Evaluating Evidence for the Role of Motor Affordances in Long-Term Memory

The two recent studies in our lab have found no evidence for a role of the motor system in long-term memory for objects and words (Pecher et al., 2015; Zeelenberg & Pecher, 2015). Because long-term memory, more than working memory, relies on semantic information and because affordances are part of conceptual knowledge we anticipated finding evidence that motor simulations support long-term memory. Our findings contrast with those of van Dam et al. (2013) who showed that performing movements after the initial study task boosted later memory for those words that were congruent with those movements (relative to words incongruent with these movements). These different effects of motor actions on memory encoding and memory consolidation are somewhat puzzling. However, as indicated, the evidence obtained with implicit memory tasks is not very strong. Clearly, additional studies are needed to delineate the conditions in which motor actions performed after initial study affect later memory.

4. FINAL CONCLUSIONS

In this paper, we have reviewed the evidence for action-based memory effects in short-term and long-term memory tasks. A meta-analysis of 11 short-term memory experiments indicated that there is no evidence for a role of motor simulations in short-term recognition and *n*-back tasks. Serial order recall tasks have provided some evidence for action-based memory effects, but the majority of these experiments used procedures that emphasized actions. Evidence that the actions associated with words and objects are automatically activated and support short-term memory is still very limited. The few experiments that have investigated action-based memory effects in long-term memory have failed to provide evidence that motor simulations play a role in memory encoding. There is limited evidence, however, that motor simulations may play a role in consolidation.

These experimental results may limit the scope of the grounded cognition framework. Although different views exist (e.g., Wilson, 2002), most accounts give a central role to the motor system, also inspired by many neuroimaging studies that show activation of the (pre)motor cortex during conceptual processing of objects or action verbs. For example, in his highly influential paper Glenberg (1997) argues that memory is for action. On his account, concepts are integrated representations of perceived and remembered actions. This view may be correct when the person is actually performing actions, or has an action goal, but perhaps less so when visual or verbal information has to be recognized or recalled. Given the highly flexible nature of our cognitive system, it may be reasonable to assume that involvement of the motor system is task dependent and as such has no central role in cognitive processing.

REFERENCES

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*, 50–68. http://dx.doi.org/10.1037/a0014411.

Apel, J. K., Cangelosi, A., Ellis, R., Goslin, J., & Fischer, M. H. (2012). Object affordance influences instruction span. *Experimental Brain Research, 233*, 199–206. http://dx.doi.org/10.1007/s00221-012-3251-0.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362–365.

Barclay, J. R., Bransford, J. D., Franks, J. J., McCarrel, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior, 13*, 471–481. http://dx.doi.org/10.1016/S0022-5371(74)80024-1.

Barsalou, L. W. (1981). Context-independent and context-dependent information in concepts. *Memory and Cognition, 10*, 82–93.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: manifestations of a compositional system of perceptual symbols. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 29–101). Hove, United Kingdom: Erlbaum.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences, 22*, 577–660. http://dx.doi.org/10.1017/S0140525X99002149.

Borghi, A. M., & Riggio, L. (2009). Sentence comprehension and simulation of object temporary, canonical and stable affordances. *Brain Research, 1253*, 117–128. http://dx.doi.org/10.1016/j.brainres.2008.11.064.

Bub, D. N., & Masson, M. E. J. (2010a). Grasping beer mugs: on the dynamics of alignment effects induced by handled objects. *Journal of Experimental Psychology: Human Perception and Performance, 36*(2), 341–358. http://dx.doi.org/10.1037/a0017606.

Bub, D. N., & Masson, M. E. J. (2010b). On the nature of hand-action representations evoked during written sentence comprehension. *Cognition, 116*, 394–408. http://dx.doi.org/10.1016/j.cognition.2010.06.001.

Bub, D. N., Masson, M. E. J., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition, 106*, 27–58. http://dx.doi.org/10.1016/j.cognition.2006.12.010.

Buccino, G., Sato, M., Cattaneo, L., Rodà, F., & Riggio, L. (2009). Broken affordances, broken objects: a TMS study. *Neuropsychologia, 47*, 3074–3078. http://dx.doi.org/10.1016/j.neuropsychologia.2009.07.003.

Casteel, M. A. (2011). The influence of motor simulations on language comprehension. *Acta Psychologica, 138*, 211–218. http://dx.doi.org/10.1016/j.actpsy.2011.06.006.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage, 12*, 478–484. http://dx.doi.org/10.1006/nimg.2000.0635.

Cimbalo, R. S., Capria, R. A., Neider, L. L., & Wilkins, M. A. C. (1977). Isolation effect: overall list facilitation in short-term memory. *Acta Psychologica, 41*, 419–432. http://dx.doi.org/10.1016/0001-6918(77)90001-4.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75–84.

Creem-Regehr, S. H., & Lee, J. N. (2005). Neural representations of graspable objects: are tools special? *Cognitive Brain Research, 22*, 457–469. http://dx.doi.org/10.1016/j.cogbrainres.2004.10.006.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

van Dam, W. O., Rueschemeyer, S.-A., Bekkering, H., & Lindemann, O. (2013). Embodied grounding of memory: toward the effects of motor execution on memory consolidation. *Quarterly Journal of Experimental Psychology, 66*, 2310–2328. http://dx.doi.org/10.1080/17470218.2013.777084.

van Dantzig, S., Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods, 43*, 145–154. http://dx.doi.org/10.3758/s13428-010-0038-8.

van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science, 32*, 579–590.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.

Downing-Doucet, F., & Guérard, K. (2014). A motor similarity effect in object memory. *Psychonomic Bulletin and Review, 21*, 1033–1040. http://dx.doi.org/10.3758/s13423-013-0570-5.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*, 671–682.

Fallon, A. B., Groves, K., & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: when CAT helps RAT, but not MAN. *International Journal of Psychology, 34*, 301–307.

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 and Review, 9*, 558–565. http://dx.doi.org/10.3758/BF03196313.

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science, 8*, 573–585. http://dx.doi.org/10.1177/1745691613498098.

Glover, S., Rosenbaum, D. A., Graham, J., & Dixon, P. (2004). Grasping the meaning of words. *Experimental Brain Research, 154*, 103–108. http://dx.doi.org/10.1007/s00221-003-1659-2.

Guérard, K., & Lagacé, S. (2014). A motor isolation effect: when object manipulability modulates recall performance. *Quarterly Journal of Experimental Psychology, 67*, 2439–2454.

Guérard, K., Hughes, W. R., & Tremblay, S. (2008). An isolation effect in serial memory for spatial information. *Quarterly Journal of Experimental Psychology, 61*, 752–762. http://dx.doi.org/10.1080/17470210701402331.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301–307. http://dx.doi.org/10.1016/S0896-6273(03)00838-9.

Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin and Review, 16*, 80–87.

den Heyer, K., Briand, K., & Dannenbring, G. L. (1983). Strategic factors in a lexical decision task: evidence for automatic and attention-driven processes. *Memory and Cognition, 11*, 374–381.

Holt, L. E., & Beilock, S. L. (2006). Expertise and its embodiment: examining the impact of sensorimotor skill expertise on the representation of action-related text. *Psychonomic Bulletin and Review, 13*, 694–701.

Jalbert, A., Saint-Aubin, J., & Tremblay, S. (2008). Visual similarity in short-term recall for where and when. *Quarterly Journal of Experimental Psychology, 61*, 353–360. http://dx.doi.org/10.1080/17470210701634537.

Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., & Doherty, S. (1989). Can you squeeze a tomato? the role of motor representations in semantic sensibility judgments. *Journal of Memory and Language, 28*, 56–77. http://dx.doi.org/10.1016/0749-596X(89)90028-4.

Lagacé, S., Downing-Doucet, F., & Guérard, K. (2013). Norms for grip agreement for 296 photographs of objects. *Behavior Research Methods, 45*, 772–781. http://dx.doi.org/10.3758/s13428-012-0283-0.

Lagacé, S., & Guérard, K. (2015). When motor congruency modulates immediate memory for objects. *Acta Psychologica, 157*, 65–73.

Lewis, M. Q. (1971). Categorized lists and cued recall. *Journal of Experimental Psychology, 87*, 129–131.

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 1—11.
Lindemann, O., Stenneken, P., van Schie, H. T., & Bekkering, H. (2006). Semantic activation in action planning. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 633—643. http://dx.doi.org/10.1037/0096-1523.32.3.633.
Mahon, B. Z. (2015a). The burden of embodied cognition. *Canadian Journal of Experimental Psychology, 69*, 172—178.
Mahon, B. Z. (2015b). What is embodied about cognition? *Language, Cognition and Neuroscience, 30*, 420—429.
Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology Paris, 102*(1—3), 59—70. http://dx.doi.org/10.1016/j.jphysparis.2008.03.004.
Marques, J. F. (2000). The "living things" impairment and the nature of semantic memory organisation: an experimental study using PI-release and semantic cues. *Cognitive Neuropsychology, 17*, 683—707.
Marques, J. F. (2006). Specialization and semantic organization: evidence for multiple-semantics linked to sensory modalities. *Memory and Cognition, 34*, 60—67.
Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology, 11*, 194—201.
Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*, 649—652. http://dx.doi.org/10.1038/379649a0.
Masson, M. E. J. (2015). Toward a deeper understanding of embodiment. *Canadian Journal of Experimental Psychology, 69*, 159—164.
Masson, M. E. J., Bub, D. N., & Breuer, A. T. (2011). Priming of reach and grasp actions by handled objects. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1470—1484. http://dx.doi.org/10.1037/a0023509.
Masson, M. E. J., Bub, D. N., & Lavelle, H. (2013). Dynamic evocation of hand action representations during sentence comprehension. *Journal of Experimental Psychology: General, 142*, 742—762. http://dx.doi.org/10.1037/a0030161.
Masson, M. E. J., Bub, D. N., & Warren, C. M. (2008). Kicking calculators: contribution of embodied representations to sentence comprehension. *Journal of Memory and Language, 59*, 256—265.
Matheson, H. E., White, N. C., & McMullen, P. A. (2014a). A test of the embodied simulation theory of object perception: potentiation of responses to artifacts and animals. *Psychological Research, 78*, 465—482. http://dx.doi.org/10.1007/s00426-013-0502-z.
Matheson, H. E., White, N. C., & McMullen, P. A. (2014b). Testing the embodied account of object naming: a concurrent motor task affects naming artifacts and animals. *Acta Psychologica, 145*, 33—43. http://dx.doi.org/10.1016/j.actpsy.2013.10.012.
McKoon, G., & Ratcliff, R. (1995). Conceptual combinations and relational contexts in free association and in priming in lexical decision. *Psychonomic Bulletin and Review, 2*, 527—533.
Mecklinger, A., Gruenewald, C., Weiskopf, N., & Doeller, C. F. (2004). Motor affordance and its role for visual working memory: evidence from fMRI studies. *Experimental Psychology, 51*, 258—269. http://dx.doi.org/10.1027/1618-3169.51.4.258.
Myung, J. Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition, 98*, 223—243.
Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226—254.
van Orden, G. C. (1987). A ROWS is a ROSE: spelling, sound and reading. *Memory and Cognition, 15*, 181—198.

Page, M. P. A. (2006). What can't functional neuroimaging tell the cognitive psychologist? *Cortex, 42*, 428–443. http://dx.doi.org/10.1016/S0010-9452(08)70375-7.

Pecher, D. (2013a). No role for motor affordances in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 2–13. http://dx.doi.org/10.1037/a0028642.

Pecher, D. (2013b). The perceptual representation of mental categories. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology*. New York: Oxford University Press.

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology, 62*, 1108–1114. http://dx.doi.org/10.1080/17470210802633255.

Pecher, D., de Klerk, R. M., Klever, L., Post, S., van Reenen, J. G., & Vonk, M. (2013). The role of affordances for working memory for objects. *Journal of Cognitive Psychology, 25*, 107–118. http://dx.doi.org/10.1080/20445911.2012.750324.

Pecher, D., Wolters, F., Stolte, F., & Zeelenberg, R. (2015). *The role of motor action in long-term memory for objects* (in preparation).

Pecher, D., Zanolie, K., & Zeelenberg, R. (2007). Verifying visual properties in sentence verification facilitates picture recognition memory. *Experimental Psychology, 54*, 173–179. http://dx.doi.org/10.1027/1618-3169.54.3.173.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science, 14*, 119–124.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: modality-specific effects of prior activation. *Psychonomic Bulletin and Review, 11*, 164–167.

Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (2002). Associative priming in a masked perceptual identification task: evidence for automatic processes. *Quarterly Journal of Experimental Psychology, 55A*, 1157–1173.

Pezzulo, G., Barca, L., Lamberti-Bocconi, A., & Borghi, A. M. (2010). When affordances climb into your mind: advantages of motor simulation in a memory task performed by novice and expert rock climbers. *Brain and Cognition, 73*, 68–73. http://dx.doi.org/10.1016/j.bandc.2010.03.002.

Poldrack, R. A. (2008). The role of fMRI in cognitive neuroscience: where do we stand? *Current Opinion in Neurobiology, 18*, 223–227. http://dx.doi.org/10.1016/j.conb.2008.07.006.

Postle, N., Ashton, R., McFarland, K., & de Zubicaray, G. I. (2013). No specific role for the manual motor system in processing the meanings of words related to the hand. *Frontiers in Human Neuroscience*, (Jan). http://dx.doi.org/10.3389/fnhum.2013.00011.

Proctor, R. W., & Miles, J. D. (2014). Does the concept of affordance add anything to explanations of stimulus–response compatibility effects? In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 60, pp. 227–266) Burlington: Academic Press.

Quak, M., Pecher, D., & Zeelenberg, R. (2014). Effects of motor congruence on visual working memory. *Attention, Perception, and Psychophysics, 76*, 2063–2070. http://dx.doi.org/10.3758/s13414-014-0654-y.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia, 47*, 388–396. http://dx.doi.org/10.1016/j.neuropsychologia.2008.09.017.

Reingold, E. M., & Goshen-Gottstein, Y. (1996). Separating consciously controlled and automatic influences in memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 397–406.

Rueschemeyer, S., Brass, M., & Friederici, A. D. (2007). Comprehending prehending: neural correlates of processing verbs with motor stems. *Journal of Cognitive Neuroscience, 19*, 855–865. http://dx.doi.org/10.1162/jocn.2007.19.5.855.

Rueschemeyer, S., Lindemann, O., Rooij, D. V., van Dam, W., & Bekkering, H. (2010). Effects of intentional motor actions on embodied language processing. *Experimental Psychology, 57*, 260–266. http://dx.doi.org/10.1027/1618-3169/a000031.

Rueschemeyer, S., van Rooij, D., Lindemann, O., Willems, R. M., & Bekkering, H. (2009). The function of words: distinct neural correlates for words denoting differently manipulable objects. *Journal of Cognitive Neuroscience, 22*, 1844–1851. http://dx.doi.org/10.1162/jocn.2009.21310.

Saccuman, M. C., Cappa, S. F., Bates, E. A., Arevalo, A., Della Rosa, P., Danna, M., et al. (2006). The impact of semantic reference on word class: an fMRI study of action and object naming. *NeuroImage, 32*, 1865–1878. http://dx.doi.org/10.1016/j.neuroimage.2006.04.179.

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex, 49*, 222–231. http://dx.doi.org/10.1016/j.cortex.2011.10.005.

Smyth, M. M., & Pendleton, L. R. (1989). Working memory for movements. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 41*, 235–250. http://dx.doi.org/10.1080/14640748908402363.

Spence, C., Nicholls, M. E. R., & Driver, J. (2000). The cost of expecting events in the wrong sensory modality. *Perception and Psychophysics, 63*, 330–336.

Taylor, L. J., Lev-Ari, S., & Zwaan, R. A. (2008). Inferences about action engage action systems. *Brain and Language, 107*, 62–67.

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. http://dx.doi.org/10.1162/0898929053124965.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 830–846.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8*, 769–800. http://dx.doi.org/10.1080/13506280042000144.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica, 116*, 185–203. http://dx.doi.org/10.1016/j.actpsy.2004.01.004.

Wickens, D. D. (1970). Encoding categories of words: an empirical approach to meaning. *Psychological Review, 77*, 1–15.

Wickens, D. D., Dalezman, R. E., & Eggemeier, F. T. (1976). Multiple encoding of word attributes in memory. *Memory and Cognition, 4*, 307–310.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: a review. *Brain and Language, 101*, 278–289. http://dx.doi.org/10.1016/j.bandl.2007.03.004.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 9*, 625–636. http://dx.doi.org/10.3758/BF03196322.

Witt, J. K., Kemmerer, D., Linkenauger, S. A., & Culham, J. (2010). A functional role for motor simulation in identifying tools. *Psychological Science, 21*, 1215–1219.

Yee, E., Chrysikou, E. G., Hoffman, E., & Thompson-Schill, S. L. (2013). Manual experience shapes object representations. *Psychological Science, 24*, 909–919. http://dx.doi.org/10.1177/0956797612464658.

Yu, A. B., Abrams, R. A., & Zacks, J. M. (2014). Limits on action priming by pictures of objects. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1861–1873. http://dx.doi.org/10.1037/a0037397.

Zeelenberg, R. (2005). Encoding specificity manipulations do affect retrieval from memory. *Acta Psychologica, 119*, 107–121. http://dx.doi.org/10.1016/j.actpsy.2004.12.001.

Zeelenberg, R., & Pecher, D. (2015a). *The effect of motor interference on long-term memory for arm and leg related action words* (in preparation).

Zeelenberg, R., & Pecher, D. (2015b). *The role of the motor system in working memory for object pictures and action words* (Manuscript submitted for publication).

Zeelenberg, R., Pecher, D., Shiffrin, R. M., & Raaijmakers, J. G. W. (2003). Semantic context effects and priming in word association. *Psychonomic Bulletin and Review, 10*, 653–660.

Zinober, J. W., Cermak, L. S., Cermak, S. A., & Dickerson, D. J. (1975). A developmental study of categorical organisation in short-term memory. *Developmental Psychology, 11*, 398–399.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science, 13*, 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. http://dx.doi.org/10.1037/0096-3445.135.1.1.

CHAPTER SIX

Understanding Central Processes: The Case against Simple Stimulus-Response Associations and for Complex Task Representation

Eliot Hazeltine*,[1] and Eric H. Schumacher§,[1]

*Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA, USA
§School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA
[1]Corresponding authors: E-mail: eliot-hazeltine@uiowa.edu; eschu@gatech.edu

Contents

Psychology of Learning and Motivation, Volume 64
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2015.09.006

Abstract

The act of choosing an action based on stimulus information and a set of arbitrary rules is termed response selection. It embodies the core of voluntary behavior and plays a critical role in most experimental tasks, yet the processes supporting it are poorly understood. Often, response selection is assumed to arise through the activation of stimulus-response (S-R) associations that bridge perceptual and motor processes. As others have pointed out before us, this conceptualization does little to account for many findings relating to choice response tasks. In the present chapter, we describe data from eight areas of research that bear on theories of response selection: task switching, the Hick—Hyman law, S-R compatibility, congruency effects, dual-task performance, task configuration, learning, and memory. We then turn to neuroimaging and neurophysiological data and examine what they indicate about how stimulus information can be flexibly mapped to motor output. Across these diverse domains, the shortcomings of the simple S-R association view consistently cohere around two related properties: First, conceptual aspects of the task trump physical properties of the stimulus and responses with regard to determining the varying demands on central processes. Second, task representations are highly structured, such that some actions are more closely related than others, and these relationships affect performance. We conclude by delineating alternative theoretical frameworks that might better capture the nature of the central processes supporting response selection.

1. INTRODUCTION

A hallmark of human behavior is flexibility. Humans can be given a novel set of instructions and perform them accurately without feedback or external reward. To do this, we must adaptively configure cognitive processes to match our goal states with the current task demands (Allport, Styles, & Hsieh, 1994; Duncan & Owen, 2000; Monsell & Driver, 2000; Norman & Shallice, 1986; Rogers & Monsell, 1995; Schneider & Shiffrin, 1977). These demands can entail a tremendous range of stimuli and responses, and the mappings between stimuli and responses can be arbitrary and dynamic. Yet, how we are able to make any voluntary response to any consciously perceived stimulus remains mysterious.

1.1 Response Selection

This operation of producing a response to a stimulus according to the current task goals is termed response selection. It is central to most voluntary behaviors and plays a critical role in most experimental tasks, but its contribution to measures of behavior is often neglected. Here we review several key findings relating to response selection and describe what they indicate about underlying cognitive processes.

Central processes like response selection are difficult to study in part because they must be distinguished from the peripheral processes that are related to stimulus identification or the specification of motor parameters. Most experimental manipulations involve changes in the stimuli or the responses and therefore likely impact peripheral processes, so it can be difficult to isolate effects on central selection processes. For example, using left and right arrows instead of colors to indicate left and right button presses may shorten the central processes that translate the stimulus to the response, but the change in stimuli will also affect perceptual processes that identify the stimuli. Because it may take more or less time to identify the direction of the arrow stimulus than to identify the color, it is difficult to attribute changes in performance to central or peripheral processes. The problem can work the other way as well: effects on central processes can also contaminate measures of peripheral processes. For example, as we argue below, learning-related improvements in performance assumed to reflect more efficient motor processing are in fact better characterized as altered response selection processes (e.g., Hazeltine, 2002). It is also unclear in some cases whether the origin of modulations in the sensitivity to stimulus information is perceptual or post-perceptual (e.g., McCann & Johnston, 1992).

A shared theme of both of the authors' collaborative and independent work is the examination of these central processes that allow individuals to take stimulus information and produce an appropriate response with minimal practice. Here, we review some findings that relate to the processes that support flexible, goal-based behavior. Our review is not exhaustive. We focus on findings that emphasize the complex nature of response selection processes along with related work from our own laboratories. Given the scope of the topic, we acknowledge that our coverage is skewed and incomplete, but our intention is to identify commonalities across a range of domains that motivate our investigations.

1.2 Investigating Central Processes

A popular approach to isolating central processes is to hold the stimuli and responses constant and manipulate the mappings between them (e.g., Duncan, 1977b; Fitts & Seeger, 1953; Hazeltine, 2005; Hommel, 1993; Huestegge & Koch, 2010; Kornblum, Hasbroucq, & Osman, 1990;

McCann & Johnston, 1992; Schumacher & D'Esposito, 2002; Simon & Rudell, 1967; Stoffels, 1996). For example, a condition in which a left arrow stimulus is mapped to a left response and right arrow stimulus is mapped to a right response can be compared to a condition in which the left arrow is mapped to the right response and the right arrow is mapped to the left response. Because the stimuli and responses are the same in both conditions, differences in response time (RT) when the mappings are changed can be attributed to central response selection processes.

A starting point for theoretical accounts of response selection is a concept borrowed from the behaviorist literature, the stimulus-response (S-R) association, which presumably bridges the perceptual and motor systems, allowing us to interact with our environments in a purposeful way. S-R associations can be instantiated in a variety of ways: they can take the form of links in a connectionist model between nodes representing stimuli and nodes representing responses (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Verguts & Notebaert, 2008) or they can take form of productions in a production system (e.g., Anderson, 1982; Anderson et al., 2004; Meyer & Kieras, 1997a,b; Salvucci & Taatgen, 2010). However, a common feature is that the presence of a particular stimulus activates a response in a feedforward fashion. That is, perceptual systems identify the stimulus, which in turn leads to the activation of the appropriate motor codes (e.g., Lien & Proctor, 2002; McCann & Johnston, 1992; Miller, 1988; Pashler, 1984).

The activation of an S-R association presumably allows stimulus information to initiate the planning and production of motor movements, but there is enduring controversy about whether, and to what extent, the relevant processes (e.g., stimulus perception, response selection, and motor preparation) work in a serial or parallel fashion. Early theories suggested that they were organized as a series of serial stages (e.g., Sternberg, 1969), so that perceptual processes completely analyzed a stimulus before central selection processes began to map the stimulus to the response and so forth. Other theories proposed that processes worked in a much more parallel fashion—with stimulus information continuously informing response processes about the probability of a likely stimulus (e.g., Miller, 1988; Usher & McClelland, 2001).

In the present paper, we embrace the latter approach and further argue that S-R associations do not provide much explanatory power for understanding many critical aspects of goal-based behavior. Specifically, we assert that accounts based on S-R associations essentially elide the complex operations that support response selection, opting instead to propose a set of stages (serial or parallel) that specify a sequence of representations that can be used to drive goal-based behavior. They do not address the coding of higher-level aspects of the tasks incorporating relationships between sets of stimuli and sets of responses—that is, they operate only on the "atomic" level of individual stimuli and individual responses. This atomic conceptualization is insufficient to explain the findings we review below.

Part of our motivation for explicating these shortcomings comes from challenges to the assumption that stimulus and response processing are distinct (e.g., Gallese, Fadiga, Fogassi, & Rizzolatti, 2002; Hommel, Müsseler, Aschersleben, & Prinz, 2001; Mechsner, Kerzel, Knoblich, & Prinz, 2001; Müsseler & Hommel, 1997; Prinz, 1990). There are multiple alternative frameworks rejecting this claim but a common thread is that the perceptual and motor processes are intimately related and temporally overlapping (e.g., Cisek & Kalaska, 2005, 2010; Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Eriksen & Schultz, 1979; Gaskell & Marslen-Wilson, 2002; McClelland, 1979; Spivey, Dale, Knoblich, & Grosjean, 2010). Therefore, instead of assuming that the flow of information is primarily unidirectional, moving from representations of stimuli to representations of responses, these accounts emphasize the bidirectional flow of information so that motor states influence perception (e.g., Klatzky, Pellegrino, McCloskey, & Doherty, 1989; Müsseler & Hommel, 1997; Proffitt, 2006).

If it is accepted that stimulus and response processing are not distinct stages of processing, then the notion of an S-R association becomes less obviously useful. That is, without the idea of distinct, serial processing stages, the need for links between a completed representation of the stimulus and a representation of the responses diminishes. Expanding the concept of the S-R association to allow for multiple, bidirectional links between individual stimulus features and responses (or components of the response) becomes tantamount to acknowledging that perceptual information is eventually transformed into motor codes—a claim that is obviously true and provides little constraint on theory—unless the links are precisely specified. In short, the use of S-R associations is often used as shorthand for central processes, much as homuncular control systems serve as placeholders in theories of executive control. However, the conceptualization of response selection as the activation of an S-R association may be more damaging; appealing to a homunculus at least implicitly acknowledges that serious theoretical questions remain, whereas the notion of an S-R association can imply that the translation process is understood, straightforward, and uninteresting.

1.3 Two Limitations of the Simple S-R Association View

In our review of the evidence against the usefulness of S-R associations for explaining voluntary behavior, we identify two primary shortcomings that may serve as categories to link findings across a range of procedures and phenomena. First, classically defined S-R associations are not rich enough to account for the patterns of transfer and flexibility of behavior reported in the literature. That is, compatibility and practice effects indicate that the encoded representations include more than just the integrated physical properties of the stimuli and responses and the mapping between them; rather, they incorporate conceptual and intentional properties of the stimuli and responses that are contingent on contextual factors. This widely recognized aspect of behavior indicates that the notion of an S-R association is too simplistic to account for the central processes that guide and select goal-based actions. We term this shortcoming the *conceptual problem*.

Second, the appropriate response to a stimulus varies according to the task and context, and the rules relating which responses are signaled by which stimuli are structured so that some rules are more closely related to others and some directly conflict. Most theorization about this aspect of behavior is based on the notion of a *task set*. Task sets are the mental representations that allow us to transform the welter of sensory information into meaningful goal-based behavior. They are often thought of as collections of individual S-R associations (e.g., Lien, Ruthruff, Remington, & Johnston, 2005; Logan, 1990; Schneider & Anderson, 2011). However, as we will argue, they include much more information that defines the boundaries of a task and facilitates learning and selection (cf., Dreisbach, Goschke, & Haider, 2007; Henson, Eckstein, Waszak, Frings, & Horner, 2014). We propose that widely studied behavioral phenomena thought to reflect changes in stimulus processing or individual S-R associations are better characterized as reflecting changes in the representation of the task and argue the current understanding for how tasks are represented and how these representations guide the motor system is lacking. This gap must be addressed if we are to develop rigorous accounts of goal-based behavior that can span the complex task demands confronting humans in real world situations. We term this shortcoming the *structural problem*.

By showing the shortcomings of theoretical accounts that rely on simple S-R associations, we hope to specify what behavioral phenomena need to be explained and to provide some initial steps based on behavioral and neural data as to how these can be addressed. We focus on eight core domains relating to the performance of choice reaction time tasks: task switching, the Hick–Hyman law, S-R compatibility, congruency effects, dual-task performance, task configuration, learning, and memory. We then examine what neuroimaging and neurophysiological studies indicate about central processes and task-set representation. We are not the first to describe these intriguingly complex aspects of response selection. Indeed, we review a broad range of work that has already established the general shortcomings of accounts relying solely on S-R associations. Our goal is to demonstrate commonalities in the findings across a range of experimental procedures and use these commonalities to delineate an approach for abandoning the reliance on the S-R association and developing an alternative theoretical framework.

2. TASK SWITCHING

Perhaps the most popular way to study how central processes are configured is with task-switching procedures. In typical task-switching studies, participants are given at least two sets of rules for how to respond to stimuli and switch between them on some successive trials. There is a considerable body of research examining the processes that underlie the activation of a task set so that a particular set of rules, which may conflict with other possible rules, determine how the participant responds to stimuli. The typical finding, especially when both tasks involve overlapping sets of stimuli, is that performance is slower on trials in which the task is switched from the previous trial compared to trials in which the task repeats (e.g., Allport et al., 1994; Jersild, 1927; Rogers & Monsell, 1995).

There are several findings in the task-switching literature that indicate that switch costs involve something other than just the activation of sets of S-R associations. For example, Logan and Bundesen (2003) argued that there is no switching of task rules in a cued task-switching procedure (but see, Brass & Cramon, 2004; Mayr & Kliegl, 2003). Instead, participants solve the task in associative manner using the cue-stimulus conjunction as a cue. Alternatively, Kleinsorge and Heuer (1999) showed that the relationship between the tasks affects the magnitude of the costs and suggested that this phenomenon reflected the hierarchical organization of the rules. They proposed that the operations involved in the switch depend on the conceptual relationships among the tasks. They concluded that the task set was best conceptualized as a multidimensional task-space as opposed to being composed of S-R associations, and switching involved selecting the appropriate control structure within this hierarchically organized space. These effects point to both conceptual (viz., the multidimensional nature of task rules) and structural (viz., the conditions under which switch costs occur) properties in task sets that are not captured by accounts based on S-R associations.

With regard to the conceptual nature of the task rules, Mayr and Bryck (2005) showed that response repetition costs can be observed even when both the stimuli are repeated but the rule used to determine which response is indicated by the stimulus is changed. That is, even on trials in which both the stimulus and the appropriate response are repeated, switch costs were observed when the underlying rule changed. Thus, the representations used by response selection processes include more than links between specific stimuli and specific responses; they appear to incorporate the underlying abstract rule linking the two together. This point was further demonstrated by the fact that the simpler horizontal switch and vertical switch rules were performed faster than the more complex clockwise and counterclockwise rules, indicating that a transformation was being applied to the stimulus to compute the appropriate response.

There is also evidence that rules, rather than simply S-R associations, can be hierarchically organized, which speaks to the richness of task-set representations. Lien and Ruthruff (2004) showed that factors that affect task-relatedness such as the timing and spatial layout of the stimuli can override task-switch costs. That is, when temporal and spatial factors encourage participants to conceptualize the sequence of responses as belonging to pairs, performing Task A followed by Task B (A-B) and then A-B again leads to more efficient performance than performing A-B then Task B followed by Task A (B-A). Thus, performing the same ensemble A-B consecutively is easier than switching the ensemble, even though AB-BA allows for the repetition of task B on the third trial. This finding suggests that it is not the loading of S-R mappings that produces the switch costs, given that the costs can be eliminated even when the new mappings must be loaded as the two element tasks are performed in rapid succession (see also, Schneider & Logan, 2007).

If abstract rules rather than S-R associations mediate response selection, then the same sets of S-R mappings might be encoded in multiple ways and indeed there is evidence to suggest this is the case. Dreisbach, Goschke, and Haider (2006, 2007) had participants learn to map eight stimuli to two response buttons. The rule-based group of participants learned the mappings based on rules involving the color of the stimulus and a judgment. The S-R group simply learned that particular stimuli required particular responses without rules. The S-R group performed faster than the rule-based group. Moreover, the rule-based group showed switch costs, whereas the S-R group did not. Thus, there is evidence that response selection can involve distinct sets of central processes with different properties. Dreisbach and Haider (2008, 2009) provided further evidence that the task representation alters central processes by showing that rules can shield response selection from irrelevant stimulus information: when the task was encoded in terms of rules, performance was less susceptible to response conflict than when the task was encoded in terms of individual correspondences.

Further evidence from task-switching experiments for the idea that responses are selected via the implementation of conceptual rules rather than fixed S-R associations comes from Schneider (2014). He had participants switch between two tasks that could both be performed on visually presented noun words (small-large and living-nonliving) that were mapped to the same pair of response keys. Critically, words were never repeated, so there was no opportunity for associations between particular words and responses to be learned (see, Waszak, Hommel, & Allport, 2003). Nonetheless, there was a robust congruency effect, as words that indicated the same response for both tasks were performed more quickly than words that indicated different responses (see also, Kiesel, Wendt, & Peters, 2007; Koch & Allport, 2006; Liefooghe, Wenke, & De Houwer, 2012; Wenke, Gaschler, & Nattkemper, 2007). Thus, this form of between-task congruency (see below) observed in task switching appears to be based on conflicting rules being applied to stimuli.

In sum, task switching costs likely derive from a range of factors, including the decoding of task cues, suppressing inappropriate mappings and reconfiguring selection processes (see, Kiesel et al., 2010). However, it is clear that the costs stem from more than simply loading sets of S-R associations in and out of working memory. Instead, switching sets appears to involve changing hierarchically organized task parameters, at least for certain types of tasks for which rules can be applied to capture relationships among multiple stimuli and responses. In such cases, the difficulty and stability of the switch depends on the relationship between tasks, and this is not easily explained when the tasks are reduced to sets of S-R associations.

Nevertheless, despite the prominence of this procedure in the study of task sets, there are other clues about the representation of task sets in domains not captured by task switching. By its nature, task switching emphasizes processes involved in inhibiting one task set and activating another, but there are revealing effects regarding the nature of central processes even when the same task set is maintained across trials. We now turn to these other behavioral phenomena and examine how S-R associations fare in these domains.

3. HICK—HYMAN LAW

The Hick—Hyman law (HHL, Hick, 1952; Hyman, 1953) is foundational to the information processing approach to cognitive psychology and describes the linear relationship between RT and the $\log_2$ of the number of S-R alternatives, provided they are all equally likely. Note that the existence of this effect alone implies that structure of task sets are important, given that adding new S-R alternatives to all of the already learned S-R alternatives would have little effect on RT. It is adding new alternatives to the limited number within the task-set that produces the lawful increases. In short, if there were no task boundaries, then the addition of any new S-R alternatives would have the same negligible effect on RT for any tasks when added to the vast array of existing S-R rules.

This aspect of the HHL is seldom emphasized but has been addressed by Schneider and Anderson (2011). According to these researchers, the HHL results from the dilution of activation of S-R alternatives as the activation is spread across more S-R alternatives. Again, as with all accounts of the HHL, the notion of boundaries between task sets is essential; some S-R alternatives belong to the current task set and receive activation and others do not, instantiating a form of task structure. Schneider and Anderson (2011) assume that the HHL derives from the memory retrieval processes and the fan effect (Anderson, 1976; Watkins & Watkins, 1975) as the task set becomes associated with more possible stimuli.

However, while the Schneider and Anderson (2011) account explains the set-specific nature of the HHL, there are other aspects of the phenomenon that are less easily accommodated by the model. Wifall, Hazeltine, and Mordkoff (2015) and others indicate that the number of responses is at least as important as the number of stimuli in determining the magnitude of the RT increase (see also, Keele, 1970; Laberge, Legrand, & Hobbie, 1969; Laberge & Tweedy, 1964; Pollack, 1959; Rabbitt, 1959). That is, adding four stimuli that are mapped to four different responses will lead to larger RT increases than adding four stimuli that are mapped to just two responses.

This finding is difficult to account for with models that assume that response selection occurs by allowing stimulus information to activate individual S-R associations in a feedforward way. Rather, it suggests that a set of response options are set up and compete for activation based on input (e.g., Usher & McClelland, 2001). It may be consistent with Duncan's *spatial transformation model* of the HHL (e.g., Duncan, 1977a,b, 1978). In this account, a set of operations are applied to the stimulus to compute the appropriate response and RTs increase with the complexity of the transformation (see also, Mayr & Bryck, 2005). If the duration of the transformation process depends on the number of responses given an arbitrary set of mappings, then this account may be able to account for why the number of responses also affects RT. However, other models of HHL include mechanisms analogous to a lookup table in which working memory is searched for the S-R association containing the observed stimulus (e.g., Hawkins, MacKay, Holley, Friedin, & Cohen, 1973; Theios, 1975). Such an approach is not sufficient to explain the role that the number of responses plays in RT. Instead, the relative importance of the number of possible responses compared to the number of possible stimuli suggests that the instantiation of a task involves the establishment of a set of response options, which compete for activation given the available stimulus information.

In sum, even the HHL, which simply relates RT to the number of S-R alternatives, suggests that response selection processes engage representations that are both structured and conceptual. They appear structured because the law relates only to S-R alternatives within the current task, not all known alternatives. The representations appear conceptual because alternatives that share a response increase RT less than alternatives involving distinct responses, consistent with abstract codes that incorporate both stimulus and response properties.

4. STIMULUS-RESPONSE COMPATIBILITY

Interactions between stimulus and response properties are perhaps most powerfully demonstrated by S-R compatibility effects, in which some pairings of stimuli with responses lead to better performance than other pairings. Like the HHL, S-R compatibility effects are some of the earliest described in the information processing tradition of cognitive psychology. Fitts and Seeger (1953) showed that RTs depended on an interaction between the stimulus set and response set, in effect demonstrating what has since become *set-level* compatibility (see, Kornblum et al., 1990). In their first experiment, three stimulus sets (viz., an array of lights in a circle, a square, or a vertical and horizontal row) were factorially crossed with three response sets (organized in a circle, a square, or a horizontal and vertical row). The two factors interacted such that the optimal response set for one stimulus was different from the optimal response set for another stimulus set (e.g., circle response organization was optimal with the circular array of lights, the square responses with the square array and the row responses with the row stimulus array). This finding was interpreted as reflecting the number of recoding steps required to translate the stimulus to a response. In their second experiment, it was demonstrated that these differences persisted across 32 sessions of practice; a finding that is difficult to account for if one assumes that direct links are formed between stimuli and responses after practice. The critical contribution here for our purposes is that the relative ease with which an S-R pair is performed depends on the relationship between the stimulus set and the response set.

A second key finding relating to S-R compatibility was reported by Fitts and Deininger (1954) the following year. They used the four different stimulus displays and eight joystick responses from Fitts and Seeger (1953) and manipulated the mapping between the stimuli and responses. Crossing the three stimulus sets with the three mappings revealed a significant interaction between the two factors, indicating that the effect of mapping was much greater for one set of stimuli than for the other sets, consistent with what is now termed *element-level* compatibility (again, see, Kornblum et al., 1990).

Accounts that rely on S-R associations do not readily provide an explanation for these forms of compatibility. The usual approach to accommodating such findings is to assume that there are additional S-R associations, established outside the experiment, that co-activate the compatible response, thereby facilitating performance when the correct response is compatible and slowing performance when it is incompatible (see, Kornblum & Lee, 1995; Lien & Proctor, 2002). However, this framework poses at least as many questions as it aims to answer. For example, why are these latent, compatible S-R associations activated only when certain combinations of stimuli and responses present in the task set? That is, the compatible association between the leftmost stimulus and leftmost response is only activated when the other possible stimuli make that stimulus the leftmost and the other possible responses make the corresponding response the leftmost. This highlights both the conceptual and structural limitation identified above. If S-R associations are invoked to explain S-R compatibility, then they must be defined conceptually and in relation to the other S-R associations within the task set but not those outside it. The same individual S-R association may be compatible or incompatible depending on the mappings of other stimuli in the task set (e.g., Duncan, 1977b, 1978).

Brain activation studies of S-R compatibility are also consistent with the idea that response selection involves more than simple S-R associations. Different regions within the prefrontal cortex have been shown to be sensitive to response selection difficulty effects like S-R compatibility depending on the stimuli (e.g., spatial or nonspatial) and responses (manual or vocal) (Nagel, Schumacher, Goebel, & D'Esposito, 2008; Schumacher, Elston, & D'Esposito, 2003). If response selection simply activated an abstract representation of a response based on an abstract representation of a stimulus, then it's not obvious why different brain regions would be involved depending on stimulus and response modality. However, if response selection involves processing more complex information, then it makes sense that different brain regions might be involved depending on the modalities of the stimuli and responses and how they are paired.

5. CONGRUENCY

The results from the studies of S-R compatibility demonstrate that the relationship between the relevant features of the stimuli and the differentiating features of the responses affects response selection, but stimulus information that is not part of the S-R mapping can also impact performance. This information can simply be irrelevant, as in studies of within-task congruency, or it can be relevant for the selection of a separate response, as in studies of between-task congruency. While this difference may seem subtle, the two literatures are largely distinct, so we will discuss each of them in turn.

5.1 Within-Task Congruency

Congruency effects are most frequently studied under conditions in which participants perform a single task and must ignore some irrelevant stimulus feature that unpredictably indicates the correct or incorrect response. There are three widely used tasks for examining within-task congruency: the Stroop task (Stroop, 1935), the Simon task (Simon, 1969; Simon & Rudell, 1967), and the flanker task (Eriksen & Eriksen, 1974). Here, we focus on the Simon task, in which participants are presented with a stimulus whose irrelevant location corresponds or does not correspond with the location of the correct response. A behavioral advantage is typically observed when the locations of the stimulus and the response correspond (i.e., a congruent trial) compared to when they do not (i.e., an incongruent trial), but several aspects of the Simon Effect indicate that it does not simply result from S-R associations between the side of the stimulus and the side of the response.

First, Wallace (1971) demonstrated that the congruency advantage depended on the location of the keys, not the anatomical effectors. Thus, when participants crossed their hands, so that the left hand was placed on the right key and the right hand was placed on the left key, responses were faster when the location of the stimulus matched the location of the key rather than the side of the effector. This finding indicates the correspondence effect does not involve motor programs but instead relates to abstract codes (see also, Hammond & Barber, 1978; Hommel, 1993). Furthermore, Guiard (1983) showed that the same physical movement (rotating a wheel) produced in response to the same stimulus could be congruent or incongruent depending on how that movement was conceptualized (e.g., as a wheel rotation vs as a hand movement).

Perhaps the most dramatic demonstration of the abstract nature of the Simon effect comes from Hedge and Marsh (1975), who showed that when S-R translation involves the reversal of relevant stimulus information, the Simon effect reverses. In their study, participants were instructed to press a red button when they saw a green stimulus and a green button when they saw a red stimulus. The buttons were in a horizontal row and the stimuli appeared to the left and right of the screen. Under these conditions, participants were faster when the stimulus occurred on the opposite side of the response. That is, reversing one relationship of the task (between the relevant stimulus color to response key color) caused the correspondence effect between another relationship of the task (between the irrelevant stimulus location and the response key location) to also reverse. This is another example of the conceptual limitation of traditional S-R association accounts. Why should reversing the S-R associations between the stimulus colors and the key colors affect the S-R associations between the stimulus locations and key locations?

These findings indicate that stimulus and response processing are intimately linked and difficult to explain with theories of central processes that assume that stimulus classification and response selection occur in discrete stages (see, Hazeltine, Akçay, & Mordkoff, 2011; Hommel, 2011; Mordkoff & Hazeltine, 2011), leading some to propose alternative models that integrate perceptual and motor processing. For example, the event coding account (Hommel et al., 2001) holds that stimulus and response information are integrated into event files, which support the production of goal-based actions (see also, Hommel, 2004). When a left stimulus must be bound into an event file with a right response, there may be conflict as the left feature of the stimulus may be incorrectly bound to the response. This approach does not provide a straightforward account of all aspects of the Simon Effect (e.g., De Jong, Liang, & Lauber, 1994; Hazeltine, Akçay, et al., 2011; Hedge & Marsh, 1975), but it does offer a foothold on how stimulus and response features appear to directly interact. Conceptual information could be integrated into event files, although there is little consideration of structure and how boundaries are instantiated between sets of S-R associations. In other words, the theory readily addresses the conceptual problem (because contextual information may also be included in the event files) but not the structural problem (because it does not explain under what situations Simon Effects should be enhanced or attenuated). Also, the mechanism as described is purely associative, linking particular features of the stimuli and responses, making it ill-suited to accommodate findings that suggest tasks are encoded as rules and can be hierarchically organized (e.g., Dreisbach et al., 2007; Hazeltine, 2005; Kleinsorge & Heuer, 1999; Mayr & Bryck, 2005). However, one might modify the account and propose something akin to "task files" that include collections of mappings so that some S-R associations are more closely related than others.

5.2 Between-Task Congruency

In addition to congruency effects between irrelevant stimulus information and the appropriate response, there can be congruency effects between ongoing operations for distinct, concurrent S-R translations, such as when we respond to one stimulus with one hand and another stimulus with the other hand. Congruency effects between concurrently performed tasks are often large compared to those associated with more traditional within-task congruency effects that rely on irrelevant information, presumably because task-relevant information activates representations more strongly than to-be-ignored stimulus features. An advantage to studying central processes by manipulating between-task congruency beyond the large magnitude of the effect is that arbitrary mappings can be used. Thus, in contrast to typical within-task congruency manipulations, researchers do not have to rely on pre-existing correspondences between stimulus features or between stimulus features and responses. The principal finding is that when people make two manual responses in close temporal proximity, the ease of responding depends on the similarity between the two movements (e.g., Kelso, Southard, & Goodman, 1979) and features of the appropriate movement for one hand can be observed in the movement of the other hand (e.g., Franz, Eliassen, Ivry, & Gazzaniga, 1996). This bimanual crosstalk phenomenon has been taken as evidence that motor codes are transmitted to the wrong effector during response execution (e.g., Heuer, 1995; Swinnen, 2002); however, some evidence suggests that the interference takes place, at least partly, at more abstract, conceptual level (see, Mechsner et al., 2001).

For example, Diedrichsen, Hazeltine, Kennerley, and Ivry (2001) examined how the stimuli used to cue the movements affected bimanual crosstalk. They found that when the movements were cued "directly" (i.e., with the presentation of stimuli at the appropriate endpoint of the movement) there was no evidence for bimanual crosstalk. The authors concluded that the cost associated with producing asymmetric movements with the two hands was associated with decoding symbolic cues into motor responses. That is, the crosstalk was located in central processes.

Hazeltine (2005) followed up on these findings with discrete button-press tasks and showed that RT depended on the relationship between the conceptual codes associated with the two responses rather than specific stimulus or response properties. The same two S-R alternatives could be compatible or incompatible, as indicated by performance measures, depending on how they were conceptualized. If participants were encouraged to think of their responses as differing in terms of distance from the body's midline, then making innermost or outermost responses with both hands was performed more quickly than making leftmost or rightmost responses with both hands. However, if participants were encouraged to think of the responses as differing along the left—right axis, then the opposite pattern of results was obtained. Note that the critical stimuli were simple crosses presented at different spatial locations and the S-R mappings for these stimuli were the same, so the online interference associated with producing two responses is not just conceptual but that the concepts themselves are determined by the organization of the task. In short, response—response congruency is not based on stimuli or responses but the conceptualization of task.

This emphasis on rules rather than individual S-R associations echoes earlier work by Duncan (1977a, 1978), who also examined how individuals simultaneously selected button presses for each hand. Duncan (1977b) presented three main empirical findings: (1) RTs for consistent mappings were shorter than RTs for inconsistent mappings; (2) there were repetition benefits when the rule repeated, even when the specific S-R association did not; and (3) errors generally reflected the application of the incorrect rule. As discussed previously, Duncan proposed the *spatial transformation model* of response selection, in which a transformation is applied to the stimulus to compute the response, but when the mapping is inconsistent a time-consuming, error-prone decision process must be invoked to determine which transformation to apply.

Duncan (1978) argued against models of response selection that produce a single S-R association, at least with tasks in which spatial transformations can be performed (see Halvorson & Hazeltine, 2015; for evidence that spatial transformation may be a special case). He noted that the idea that selection relies on rules rather than individual S-R associations is not new, citing Welford (1958), Shaffer (1965), and Rabbitt and Vyas (1973), but that most accounts of response selection emphasize S-R associations.

In sum, along with findings from experiments examining within-task congruency, studies of between-task congruency indicate that the selection of responses involves the activation of representations that include more than just stimulus and response information. Rather, abstract rules appear to play a critical role in determining how relevant and irrelevant information impinges on response selection. The notion of an S-R association provides little insight into how the congruency effects described here might emerge.

6. DUAL-TASK PERFORMANCE

While the studies of compatibility and congruency examine how activated stimulus and response information affects the selection of a single response, dual-task studies examine how ongoing processes for temporally overlapping tasks interfere with each other. In essence, this domain is closely related to the between-task congruency work, although the relationship is seldom explored (but see, Hazeltine, Diedrichsen, Kennerley, & Ivry, 2003; Hazeltine, Teague, & Ivry, 2002; Huestegge & Koch, 2010; Ivry, Franz, Kingstone, & Johnston, 1998; Ivry & Hazeltine, 2000; Navon & Miller, 1987). Dual-task experiments tend to use tasks that appear unrelated (e.g., responding manually to colored shapes and vocally to tones) and to not consider differences between specific combinations of stimuli across the two tasks. A central finding in dual-task research is that performing two tasks in an overlapping fashion almost always produces performance costs on at least one of the tasks, even when they involve distinct stimulus and response modalities.

6.1 Single-Channel Models

The dominant theoretical accounts of this cost posit a single, central processor, with access to all sensory and motor systems, that intervenes between stimulus processing and response production (e.g., Anderson et al., 2004; Anderson, Taatgen, & Byrne, 2005; Dux, Ivanoff, Asplund, & Marois, 2006; Lien & Proctor, 2002; McCann & Johnston, 1992; Pashler, 1994b; Pashler & Johnston, 1989). The idea is that only a single S-R association can be activated at a time so when two stimuli are categorized at nearly the same time, the next processing stage, response selection, must be delayed for one of the categorized stimuli, and a dual-task cost is observed. However, it should be emphasized that the assertion that response selection is mediated by S-R associations and the assertion that response selection is limited to a single task at a time are independent, and accepting one claim requires no commitment to the other. In short, the single-channel account holds that there is a bottleneck at response selection; response selection can only be engaged for a single task at a time, so whenever it is simultaneously required by two tasks, processing for one of them is deferred. What this has to do with S-R associations we address later in this section.

The single-channel account addresses some of the vexing problems faced by models of response selection. By assuming that there is a unified central processor, such accounts can explain how we are able to map any stimulus to any response. Without a single, central response selection mechanism, it is difficult to envision, within the S-R association framework, how a range of perceptual codes can interface with a range of response systems, and how processing is controlled (although the proposition that response selection is restricted to one task at a time can be accepted without positing a single processor that selects all responses; see, Pashler, 1994a). Moreover, the notion of a single processor accounts for why dual-task costs are observed regardless of the modalities of the stimuli and responses (Pashler, 1989, 1990; Smith, 1967; Welford, 1952, 1967), and phenomenologically, it fits with our sense that we are able to think of only a single thing at any given moment.

However, this approach may introduce as many problems as it solves. How does a single mechanism operate on such diverse sets of inputs and outputs flexibly and with little or no practice? Can such a set of processes spanning so many inputs and outputs be meaningfully described and investigated as a single system? And how does this account accommodate findings that suggest irrelevant stimulus information is translated into response codes in parallel with the relevant stimulus information (see, Eriksen & Schultz, 1979; Hommel, 1998; Lien & Proctor, 2002)?

Neuroimaging studies have been brought to bear on whether a single mechanism subserves response selection and the data are mixed. Some neuroimaging studies directly investigate dual-task performance by comparing single-task conditions to dual-task conditions (e.g., Klingberg, 1998; Szameitat, Schubert, Müller, & von Cramon, 2002) or conditions in which the tasks are separated by a long interval to conditions in which the tasks are separated by a short interval (e.g., Dux et al., 2006; Herath, Klingberg, Young, Amunts, & Roland, 2001). The typical finding in such experiments is that regions in the prefrontal cortex, often within the inferior frontal gyrus, are more active when two tasks must be performed close together in time (but see, Dux et al., 2006; Jiang, Saxe, & Kanwisher, 2004). However, these results may have as much to do with executive control processes that coordinate task performance than with the central processes that mediate response selection itself (see, Buss, Wifall, Hazeltine, & Spencer, 2014; Dux et al., 2006).

An alternative approach that avoids this interpretive issue is to manipulate the duration of central processes and determine whether the neural regions sensitive to this manipulation depend on the types of stimuli and responses used (for a hybrid approach, see Stelzel, Schumacher, Schubert, & D'Esposito, 2006). This procedure has often produced results that suggest that neural regions associated with central operations are contingent on the modalities of the stimuli (e.g., Schumacher & D'Esposito, 2002; Schumacher et al., 2003), although there have been findings suggesting a single processor that operates across multiple domains (e.g., Jiang & Kanwisher, 2003a,b). For example, in a neuroimaging study examining congruency effects, Schumacher, Schwarb, Lightman, and Hazeltine (2011) observed that the differences between incongruent and congruent trials depended on the modality of the stimuli. That is, incongruent auditory temporal flankers (i.e., the flankers preceded the target) increased activation compared to congruent auditory flankers in a set of regions that was mostly distinct from those that were sensitive to whether visual temporal flankers were congruent or not. In short, the evidence from the neuroimaging literature that response selection engages a single, common processor is weak.

Behavioral studies have sought to test the assumption that dual-task performance is limited by an immutable response selection bottleneck shared by unrelated tasks. There are many results consistent with this idea (e.g., McCann & Johnston, 1992; Pashler, 1984, 1994a; Pashler & Johnston, 1989). However, support for a response selection bottleneck has been undermined by work investigating dual-task costs with moderate practice (Hazeltine et al., 2002; Schumacher, Seymour, Glass, Kieras, & Meyer, 2001). These studies paired visual stimuli with manual responses and auditory stimuli with vocal responses. Both studies reported the disappearance (or substantial reduction) in dual-task costs with increased practice. This may suggest that dual-task interference is caused not by a bottleneck in response selection but by strategic factors or crosstalk between codes associated with the two tasks. If this is the case, then it is another indication of both the conceptual and structural limitation in the simple S-R association account. However, there are alternate interpretations of these results that attempt to save the idea of a bottleneck in response selection (e.g., Anderson et al., 2005; Dux et al., 2009; Ruthruff, Johnston, & Van Selst, 2001) so these data are not dispositive for the nature of response selection.

While much of the behavioral work on dual-task performance has focused on visual-manual tasks paired with auditory-vocal tasks to avoid peripheral interference, there is evidence that the stimulus and response modalities do matter, outside of their indirect effects on task difficulty. Hazeltine, Ruthruff, and Remington (2006) compared dual-task performance on two task pairings (see also, Hazeltine & Ruthruff, 2006; Hazeltine & Wifall, 2011). One group of participants performed a visual-manual task (i.e., visually presented words mapped to manual button presses) and an auditory-vocal task (i.e., auditory tones mapped to spoken words) and another group of participants performed a visual-vocal task (i.e., visual words mapped to spoken words) and an auditory-manual task (i.e., tones mapped to button presses). Dual-task costs were much larger for the participants performing the visual-vocal and auditory-manual tasks, even when single-task RTs were equated, a finding that is particularly striking when one considers that on dual-task trials both groups of participants are seeing a word and hearing a tone, and pressing a button and saying a word. This finding indicates that interactions between ongoing central processes for two tasks depend on the task structure, not just the individual stimuli and responses. The difference in costs persisted across 16 sessions of practice, which is difficult to explain if one assumes that selection is mediated by S-R associations (an example of the conceptual limitation). Why should some associations interfere with each other more than others, given that the stimuli and responses were highly similar across the two conditions, unless contextual information was also encoded in the task representation?

6.2 Crosstalk

One explanation for these modality-dependent effects is crosstalk (i.e., interference between S-R associations for each task). That is, the amount of crosstalk between the two tasks may depend on their composition. Evidence for the idea that crosstalk provides a critical limitation on dual-task performance, rather than a unitary response selection mechanism, was reported by Halvorson, Ebner, and Hazeltine (2013), who investigated the near absence of dual-task costs across various combinations of tasks. The researchers were examining the claim that dual-task costs are dramatically reduced when tasks are ideomotor-compatible. The term ideomotor-compatible refers to tasks in which the relevant stimulus is highly similar to the consequences of the response. For example, saying the word "dog" to the auditory presentation of the word "dog" is ideomotor-compatible because the sensory consequences of saying the word "dog" are highly similar to the auditory stimulus "dog." Greenwald (1970, 1972) proposed that such tasks do not require the central processes that are engaged by other tasks and thus are not subject to the same capacity limitations. However, Halvorson, Ebner et al. (2013) showed that a single ideomotor-compatible task was not sufficient to nearly eliminate dual-task costs. Thus, the authors argued that dual-task costs do not appear to be avoided because particular S-R associations do not require central processes. Rather, dual-task costs are avoided because the central processes for the two tasks involve codes that are not confusable (see also, Navon & Miller, 1987).

This proposal was further tested by Halvorson and Hazeltine (2015), who pitted two accounts of why pairs of ideomotor tasks do not produce dual-task costs. According to the ideomotor account, tasks in which the stimulus overlaps with the environmental consequences of the response bypass response selection processes so that there are no capacity limitations. According to the crosstalk account, dual-task costs result from crosstalk between simultaneously active codes associated with S-R translation. When a purely spatial task is paired with a purely verbal task, there is no crosstalk between the codes and no costs are observed. To determine which explanation provided the better account of dual-task costs, the researchers reversed the mapping of an ideomotor task so that there was still a purely spatial task paired with a purely verbal task, but the environmental consequences of the responses no longer overlapped with the corresponding stimulus cues. These new task pairings produced no dual-task costs, indicating that ideomotor compatibility was not necessary to eliminate costs and that crosstalk was the more likely source. In this way, the findings indicate that the specific content of the central processes determines the magnitude of the dual-task interference, not just the generic activation of S-R associations (see also, Hazeltine et al., 2006). Under some circumstances, responses can be selected simultaneously with little cost.

The crosstalk account may be related to the between-task congruency effects described above. However, unlike dual-task costs, which are observed regardless of the particular stimuli that appear on the two tasks, between-task congruency effects do not involve specific combinations of stimuli/responses across the two tasks. Nonetheless, the underlying mechanism between the two phenomena may be the same. In other words, executive control may delay central processes on one task, resulting in generic dual-task costs, so that item-specific interactions between central processes do not occur (cf., Meyer & Kieras, 1997a,b). Alternatively, each of the items for one of the tasks may produce similar interference with each of the items for the other task, even though the magnitude of the interference is determined by the particular combination of items on each task; the items for one task would produce different amounts of interference when paired with the items for another task.

In any case, dual-task studies indicate that interference between concurrently performed tasks is content-specific (see, Hazeltine et al., 2006). That is, while costs are difficult to avoid, their magnitude depends on the relationship between the tasks. Such a state of affairs is not entirely inconsistent with models of response selection based on S-R associations, but there is presently no principled explanation for why any two particular S-R associations should interfere with each other more than any other pair. That different costs can be observed depending on how the same two stimuli are paired with the same two responses indicates that central codes containing more than just stimulus and response information must be considered in accounts of response selection.

7. TASK CONFIGURATION

Studies of dual-task performance indicate that behavioral costs depend on the structure of the tasks rather than simply the stimuli that need to be identified and the responses that need to be produced. The importance of task structure is also apparent in studies examining how the selection of a response on one trial affects response selection on a subsequent trial. That is, while dual-task studies investigate how concurrently performed central processes interact, studies of task configuration probe how the selection of a response at one point in time affects the selection of responses at future points in time. This description suggests that the behavioral effects examined in studies of task configuration may be learning phenomena, and indeed task configuration and learning may be closely related (see, Botvinick, 2007; Verguts & Notebaert, 2008), but for the present purposes, we make a distinction between effects that stem from the performance of a single, identifiable trial, as is typically the case in studies of task configuration, and effects that stem from the performance of multiple, possibly heterogeneous trials, as in most studies of practice and learning.

7.1 Cross-Trial Control

Changes in the configuration of task operations are the purview of cognitive control processes. For example, it is widely assumed that cognitive control is invoked to limit the influence of task-irrelevant information (e.g., Botvinick et al., 2001; Duncan, 2001; Egner & Hirsch, 2005; Logan & Gordon, 2001; Norman & Shallice, 1986), and evidence for the dynamic nature of these control processes can be found in studies demonstrating that resolving conflict on one trial can affect sensitivity to conflict on the subsequent trial. Gratton, Coles, and Donchin (1992) examined the congruency effect in the flanker task and found that its magnitude on one trial depended on the congruency of the previous trial. On trials in which the immediately previous trial was congruent, the flanker effect was larger than on trials in which the immediately previous trial was incongruent. This basic phenomenon is observed across a range of tasks used to study response competition, including the flanker task (e.g., Akçay & Hazeltine, 2007), the Simon task (e.g., Hazeltine, Akçay, et al., 2011) and the Stroop task (e.g., Freitas, Bahar, Yang, & Banai, 2007), and has been called conflict adaptation (e.g., Ullsperger, Bylsma, & Botvinick, 2005), the Gratton Effect (e.g., Notebaert & Verguts, 2008), sequential modulations (e.g., Hazeltine, Lightman, Schwarb, & Schumacher, 2011) and the congruency sequence effect (CSE, e.g., Schmidt & Weissman, 2014).

Because this phenomenon, which we will call CSE, presumably reflects changes in response selection processes, it can be used to probe their structure. Thus, there is extensive debate regarding the boundaries of the

CSE; that is, researchers have examined what features consecutive trials need to share in order for the congruency of one trial to affect the magnitude of the congruency effect on the next? Much of the existing work on this question has focused on the role of stimulus properties. For example, Funes, Lupiáñez, and Humphreys (2010) observed that the CSE did not occur when consecutive trials involved Simon and flanker conflict or vice versa (see also, Blais, Robidoux, Risko, & Besner, 2007; Egner, Delano, & Hirsch, 2007; Egner & Hirsch, 2005; Notebaert & Verguts, 2008), and therefore concluded that control processes operated on specific stimulus dimensions (but see, Fernandez-Duque & Knight, 2008; Freitas et al., 2007).

However, as with the studies of compatibility and congruency, there is evidence that the source of the effect is more abstract and conceptual than physical stimulus features or dimensions. Hazeltine et al. (2011) used a temporal flanker task (i.e., the flankers preceded the target) in which stimuli were presented either visually or aurally and observed that, under the appropriate conditions, CSE were observed when the stimuli changed from one modality to another. Because shared stimulus features appeared to play little role in whether the congruency of one trial affected the magnitude of the congruency effect on the next, the researchers proposed that CSE are not constrained by hard boundaries based on the organization of the perceptuomotor system but rather are determined by the individual's representation of the task; CSEs depend on the extent to which consecutive trials belong to the same task. In short, Hazeltine and colleagues proposed the patterns of CSE reflected the structure of central processes rather than perceptual or attentional mechanisms (see also, Akçay & Hazeltine, 2008, 2011; Hazeltine, Akçay, et al., 2011). This is similar to an episodic account proposed by Spapé and Hommel (2008), which holds that control settings are retrieved based on the episodic context. However, the task-set representation account emphasizes that factors beyond stimulus information can affect the magnitude of the CSE. That is, the similarity of consecutive trials relative to the similarity of other possible trials might affect the CSE, not just whether individual stimulus or response features overlap. Thus, the absence of CSE may reflect the active representation of distinct sets rather than a failure of retrieval processes. There is an obvious parallel here to the congruency and dual-task work described above: interactions between various sources of task-related information appear to incorporate aspects of the task goals and task structure, not just surface features of the stimuli and responses.

7.2 Partial Task Precuing

A second form of control relevant to the question of how task representations are structured can be observed in precuing tasks in which participants are given partial information about the upcoming stimulus that will signal a choice response. This situation, in which some aspects of the upcoming stimulus or response are cued, differs from the control studies described above in that the initial stimulus that influences subsequent performance is part of the same trial, does not require an overt response, and is (explicitly) informative. Nonetheless, the results of such studies can reveal much about the structure of central processes.

Rosenbaum (1980, 1983) demonstrated that when the possible responses were signaled by stimuli specifying particular movement parameters (e.g., which hand was to make the movement or the direction of the movement), not all precues provided the same benefits in performance, even when they provided equivalent reductions in the number of possible stimuli and responses (but see, Goodman & Kelso, 1980). For example, a precue indicating that the appropriate response would be one of the four involving the left hand produced greater reductions in RT than a precue indicating that it would be one of the four upward movements. Such an apparent violation of HHL (see Section 3) would not be obtained if the underlying selection process was simply an unorganized set of S-R associations a subset of which were primed by the precue. Thus, this result is another example of the boundaries between various components of the task representation; more information must be encoded into the task than simple S-R associations.

Rosenbaum attributed this effect to the organization of the motor system (see also, Miller, 1982). However, Reeve and Proctor (1984) showed that the configuration of the hands partly determined what precue information produced the largest decrements in RT, suggesting that the precue shortened central processes involved in response selection rather than peripheral motor programming processes. The debate persisted (e.g., Miller, 1985; Reeve & Proctor, 1985, 1990), in part because different patterns of advantages were obtained depending on the stimuli used and the configuration of the hands and response buttons. Adam, Hommel, and Umilta (2003) manipulated the locations of the stimuli and locations of the responses independently and showed that these two factors along with the type of precue produced a three-way interaction. Critically, the type of precue that was most effective (i.e., produced the largest decrements in RT) depended on the relationship between the stimuli and the responses. These findings indicate that the ability of individuals to use precues to prepare upcoming responses depends on the particular S-R mapping rather than organization of the stimuli or the motor system. Therefore, precuing effects reflect central processes.

In sum, these studies of cognitive control indicate that task set representation plays a critical role in performance. The selection of responses based on stimulus information affects the immediately subsequent selection of response in a way that depends on the relationships of both the stimuli and responses of the two selections. This indicates that the relevant control processes are operating on representations that include both stimulus and response information. Similarly, when a cue indicates that a subset of the possible stimuli will be presented, the reduction in RT is contingent on structure of the task. Conceiving of tasks as collections of S-R associations provides little headway for explaining how control processes change central processes in anticipation of upcoming events.

8. LEARNING AND PRACTICE

Dual-task studies indicate that task representation plays a critical role in determining the magnitude of dual-task costs; that is, they govern the way concurrently performed tasks interact. The same is true of the task configuration studies, although these describe interactions between events separated by short intervals of time. However, aggregations of trials can affect the performance of subsequent trials across larger timescales. Although we have emphasized the ability for humans to perform arbitrary S-R mappings with minimal practice, well-learned behaviors, which are often described as becoming effortless because they are encoded as automatic S-R associations (e.g., Hommel, 2000; Lien & Proctor, 2002; Logan, 1988; Schneider, 1985), are actually complex and manifold.

8.1 Nonhuman Animal Learning

An obvious starting point for a discussion of the learning literature is conditioning. Rescorla's (1988b) review of the conditioning literature emphasizes that conditioning is generally mischaracterized as the formation of associations between stimuli and responses when in fact it is something much more complex (see also, Rescorla, 1988a; Rescorla & Wagner, 1972). In short, instead of establishing links between stimuli and responses, conditioning might be better described as reconciling the organism's internal model of the external world with the true state of the external world.

Rescorla (1988b) identifies several key findings to illustrate this point. First and most fundamentally, a conditioned stimulus will only elicit a particular response under certain circumstances; it is not the case that a stimulus always automatically produces a learned response. For example, an animal will not press a button to receive food when a light is presented if the animal is not hungry. This basic point illustrates that the animal hasn't learned an automatic association between the light and the button press but instead has learned a set of relationships among the light, the lever, and the outcome. Second, it is not just the co-occurrence of the unconditioned stimulus and the conditioned stimulus that leads to learning, but how predictive the unconditioned stimulus is of the conditioned stimulus; the base-rate of the unconditioned stimulus is critical. Moreover, stimuli can also be prevented from forming associations with an unconditioned stimulus if associations already exist between that unconditioned stimulus and another stimulus, a phenomenon known as blocking (Kamin, 1968). For example, an animal that has been trained to associate a light with food will salivate to the light. But if the light and a tone are then paired together before food is presented, the animal will not learn to salivate to the tone—even if it is as predictive of food as the light is.

Third, different conditioned stimuli will produce dramatically different responses, even when paired with the same unconditioned stimulus. An example of this is the Garcia Effect, which describes the fact that some conditioned stimuli (e.g., tastes) are more easily associated with particular unconditioned stimuli (e.g., sickness) than other conditioned stimuli (e.g., lights and sickness; Garcia & Koelling, 1966). Similarly, Rescorla (1988b) points out that different conditioned stimuli will evoke different behaviors when paired with the same unconditioned stimuli. For example, visual stimuli will typically elicit pecking behavior in pigeons when paired with food, whereas a diffuse tone will elicit increases in general activity (see also, Pinel & Treit, 1979).

These differences should not imply that the physical properties of the stimuli are solely responsible for driving learning. It has also been shown that competition between stimulus categories, rather than the individual stimuli, can produce overshadowing (Soto & Wasserman, 2012). That is, items that provide redundant information will interfere with category learning if they belong to the same category but not if they belong to varied categories. These findings point to another conceptual limitation in the simple S-R association account of learning. They indicate that the learning is more abstract that simply pairing a particular stimulus with a particular response. Instead, they suggest the animal is building a complex model of the world that can be used for goal-based behavior. Finally, second-order conditioning (in which stimuli gain the ability to reinforce behavior through prior conditioning; see, Rescorla, 1972) also indicates the complexity of what is learned during conditioning.

While the above phenomena emphasize the conceptual nature of conditioning, there is also evidence for hierarchical structure within conditioned learning. Rescorla (1988b) notes that compound stimuli can be represented as distinct from either of component stimuli. Second-order conditioning procedures show that animals are able to represent compound stimuli (e.g., red horizontal bars) and use them as associates distinct from either component alone (e.g., red bars or horizontal bars). Animals are also able to use "occasion-setting" stimuli to cue positive or negative relations between other stimuli rather than to cue the occurrence of a particular stimulus (Holland, 1983). In this sense, the occasion-setting stimulus acts much like a task cue, indicating relationships among other stimuli that are independent of the relationships between those other stimuli and the occasion-setting stimulus itself. Thus, a stimulus can signal relationships between other stimuli independently of its associations with these other stimuli, indicating that the associations appear to be hierarchically organized and not restricted to linking individual events.

8.2 Human Learning and Practice

Studies of human learning have borrowed the concept of the S-R association from the animal learning literature to explain how stimulus information is efficiently transformed into motor activity. However, if the notion of a simple S-R association is insufficient to capture nonhuman animal behavior in constrained conditioning procedures, it would seem to provide a poor foundation for more complex human behaviors. There are examples in the cognitive literature that make this point explicitly. Mayr and Bryck (2005), for instance, demonstrated that the learning of S-R alternatives includes more than just the particular stimulus and particular response. In their experiments, the frequency of the S-R pairs was held constant but the S-R rules varied. The more frequently applied S-R rules showed greater learning than less frequently applied S-R rules, indicating that the abstract rule was incorporated in the sequence or that the learning was embedded in a rule-specific process.

There are parallels to the Mayr and Bryck (2005) findings in the serial reaction time (SRT) task literature. The SRT task is a popular means of examining motor learning (e.g., Berns, Cohen, & Mintun, 1997; Doyon et al., 1997; Frensch, Buchner, & Lin, 1994; Grafton, Hazeltine, & Ivry, 1995; Hazeltine, Grafton, & Ivry, 1997; Hazeltine & Ivry, 2002; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Nissen & Bullemer, 1987; Rauch et al., 1995; Schumacher & Schwarb, 2009; Schwarb & Schumacher, 2012; Seidler et al., 2002; Toni, Krams, Turner, & Passingham, 1998; Willingham, Nissen, & Bullemer, 1989), and several studies indicate that learning involves codes that are more than just stimuli and responses. Schwarb and Schumacher (2010) showed that learning depends on the relationship between the S-R alternatives (i.e., involves response selection; also see, Schwarb & Schumacher, 2009) and not on the specific features in the stimuli and responses. Willingham and colleagues (Willingham, 1998; Willingham, Wells, Farrell, & Stemwedel, 2000) proposed that learning in the SRT task was based on abstract locations that could be accessed by distinct effectors and scaled according to the task demands (see also, Grafton, Hazeltine, & Ivry, 1998; Keele, Jennings, Jones, Caulton, & Cohen, 1995; Wright, 1990). Hazeltine (2002) examined how practicing a motor sequence would transfer to novel sets of stimuli and novel responses. In this study, responses evoked specific tones, and transfer was best when the practiced and novel movements produced the same sequence of tones, even when both the stimuli and responses were novel. Hazeltine (2002) concluded that learned sequence representations included distal effects (see also, Elsner & Hommel, 2001; Ziessler & Nattkemper, 2001, 2002), consistent with the claim that goal states, rather than just S-R associations, are part of the task set representation.

The transfer studies indicate that learning in the SRT task can include conceptual information, but there is also evidence that task structure can shape learning. Schumacher and Schwarb (2009) showed that sequences of responses were better learned when the interval between the sequenced task and the random, distractor task was increased. This finding may reflect that separating the tasks in time may eliminate interference in central response selection processes allowing for associations to be formed among the separate events. It is also possible that the longer interval provides a cue to subjectively organize the tasks. Halvorson, Wagschal, and Hazeltine (2013) used an instructional manipulation in which two groups of participants were encouraged to either consider two alternating sets of S-R mappings as belonging to the same task or different tasks. After a pretraining period that differed for the two groups, a training phase that was identical for both groups was performed in which the stimuli for one of the S-R sets followed a repeating sequence. The group that was encouraged to conceptualize their actions as belonging to two separate tasks showed significantly more sequence learning than the group that conceptualized the two S-R sets as related. The authors argued that the task representation influenced the encoding of relationships between stimuli and could serve to protect learning from random, irrelevant stimuli.

Along these lines, Freedberg, Wagschal, and Hazeltine (2014) used a chord-learning task in which participants made two simultaneous responses and showed that the conceptual-relatedness of the stimuli and responses, rather than their modalities, determined whether practiced combinations were performed more quickly than novel combinations after practice. In other words, simultaneous responses are coded as a single action only when they were conceptually related. This finding points to both the conceptual and structural limitations of the S-R association account response selection. Not only is conceptual information encoded in the task set representation, but that information affects whether responses are encoded as related or unrelated.

While these studies demonstrate that task representations may affect the way that discrete actions are linked together, there is also evidence that motor adaptation may be sensitive to the way the task is conceptualized. Taylor and Ivry (2013) examined how motor adaptation to rightward movements transferred to leftward movements and discovered that the configuration of target locations determined whether the adaptation was rotational or translational, not whether the perturbation was rotational or translational. That is, participants learned a rotational adaptation if the target locations were arranged in a circle even if the perturbation was translational. The authors concluded that adaptation was not confined to low-level representations of the movement; instead, it appears that the task representation plays a critical role.

It has been proposed that as a task becomes highly practiced, behavior ceases to be driven by more complex verbal codes and instead is based on more automatically activated S-R associations (e.g., Anderson, 1982; Logan, 1988; Pashler & Baylis, 1991). However, while there is evidence central processes become more efficient with practice (e.g., Anderson et al., 2005; Dux et al., 2009; Reisberg, Baron, & Kemler, 1980; Ruthruff, Johnston, Van Selst, Whitsell, & Remington, 2003), there is little evidence for the formation of direct links between stimulus and response codes. Rather, across a wide array of task procedures, complex contextual information appears to be encoded into the task set representation, and these contextual effects lead to a diverse set of behavioral consequences. In sum, learning occurs at multiple levels of representation and captures a diverse array of relationships between task components. Nonetheless, even in cases of nonhuman animal learning and motor adaptation, there is evidence that the encoded information includes more than simple S-R associations.

9. MEMORY

Thus far, we have outlined the conceptual and structural limitations in the traditional S-R association account of response selection. We have shown how this account fails to explain a wide range of fundamental effects relating to human and animal performance. Another area where the S-R association account also fails to account for the data is memory retrieval.

9.1 Priming

There may not be a more direct case for S-R associations than priming. Priming is thought to occur directly from spreading activation across associations in memory (e.g., Logan, 1990; Meyer & Schvaneveldt, 1971). However, even in this case, the data suggest that more than S-R associations matter. Priming has been shown to occur at an abstract conceptual level that is superordinate to the stimuli and responses used in the task. For example, Horner and Henson (2009) found reduced priming to stimuli in a repetition priming procedure when the classification of the stimuli changed (e.g., bigger to smaller and vice versa) from one trial to the next and also when the response changed from vocal to manual. That is, the same stimuli and response systems were used but priming depended on the higher-order associations between the S-R pairs. These results suggest that priming does not only activate a stimulus representation, but that aspects of the entire task set are primed—and when aspects of the task set change, priming is reduced. As reviewed by Henson et al. (2014), priming occurs across many dimensions of the S-R pair. They propose that "SR bindings are more than simple associations between a specific concept and motor act: they are complex, structured representations that simultaneously bind multiple levels of stimulus, response and task representation (p. 382)." This characterization is clearly in line with our conclusions from the congruency, dual-task, control, and learning literature and could serve as a succinct description of the task sets that support central processes.

9.2 Explicit Memory

There are other possible effects of task sets in the memory literature as well. Consider release from proactive interference (Wickens, Born, & Allen, 1963), which is observed when participants attempt to store a number of items in working memory across a series of trials. The items come from the same category (e.g., animals) for a few trials then change to a new category (e.g., furniture). Participants' memory performance decays across trials within a category (due to proactive interference; Keppel & Underwood, 1962) but improves after a category switch. There are a number of proposed mechanisms to explain this release from proactive interference. For example, the change in category may affect attention, memory encoding or some combination of factors (Kintsch, 1970; Wickens, 1970, 1972). Any of these proposed mechanisms may be a consequence of a shift in task set. If participants develop a task set for the procedure that includes the category domain, then a shift in category may lead to the creation of a new task set. Consequently, the attention, memory and control processes working within the original task set no longer affect performance under the new set.

Task set representation may also explain the importance of *retrieval mode*. Retrieval mode is a hypothesized cognitive state in which people are oriented towards remembering existing knowledge (Tulving, 1983). This state improves the effectiveness of currently present stimuli to act as retrieval cues. Retrieval mode involves activating a task set for memory retrieval, so attention and memory processes are allocated to the appropriate stimuli and the associates in memory of those stimuli. The connection between retrieval mode and task set is amplified by findings showing the benefits of a related process of *retrieval orientation*. Retrieval orientation is, in essence, the task set adopted as individuals interpret retrieval cues (e.g., phonologically or spatially, Rugg, Allan, & Birch, 2000; Wilding, 1999). When these cues are interpreted appropriately, retrieval performance is better (for a review see, Rugg & Wilding, 2000). Thus, retrieval mode and orientation affect stimulus processing and change memory retrieval efficiency in much the same way that task sets organize and segregate response selection.

10. SUMMARY OF THE BEHAVIORAL PHENOMENA

We have described a range of behavioral effects observed across disparate procedures and a common theme has emerged: Response selection relies on complex representations that incorporate more than simple S-R associations. The task switching literature indicates that in many cases tasks appear to be represented as rules rather than sets of associations. Studies probing the source of the linear relationship between the $\log_2$ of the number of S-R alternatives and RT, known as Hick—Hyman law, show that these alternatives are partitioned into task sets, so that only a very few of them impinge on response selection processes at a given time. Furthermore, stimuli associated with the same response can produce smaller increases in RT than stimuli associated with different responses, suggesting that the perceptual system does not simply activate an individual S-R association in a feedforward manner. The surprising complexity of S-R compatibility phenomena also emphasizes that S-R alternatives must be grouped and coded in a relational fashion. The S-R compatibility of a given alternative is determined in comparison to other possible alternatives rather than in terms of absolute stimulus and response properties. Both within-task and between-task congruency effects are consistent with the proposal that response selection processes employ representations that prioritize conceptual and relational information. The congruency of irrelevant stimulus or response information depends on its conceptual relationship to the selected action, not the irrelevant stimulus' or response's physical properties.

More complex behavioral phenomena involving control processes to coordinate task operations reinforce the need for task representations that are abstract and structured. When two tasks are performed concurrently, the magnitude of the dual-task interference depends on the structure of the tasks, not just the stimuli and responses presented and produced on a given trial. The importance of structure is also apparent in studies of task configuration. The ability to alter task operations based on changing expectations about the upcoming stimuli depends on the relationships among the S-R mappings, not just the relationships among stimuli or among responses. With regard to cross-trial cognitive control, this means that task representations determine whether the congruency of the previous trial modulates the magnitude of the congruency effect on the current trial. In other words, the modulation depends not just on the stimuli and responses of the two trials, but on whether the two trials are conceptualized as belonging to the same task. With regard to partial-task precuing, this means that the benefit of a precue depends not just on how the subset of stimuli indicated by the precue corresponds with the perceptual groupings of the stimuli or motor parameters of the corresponding responses, but also on an interaction between these two factors—that is, the benefit depends on how the information provided by the precue aligns with the structure of the task.

Finally, the importance of task structure can be observed at larger time scales, as it guides the learning of contingencies between simultaneous or consecutively performed actions. Note that in the nonhuman animal learning literature, there is strong evidence that learning is based on minimizing prediction error, not the formation of associations between stimuli and responses, and that these predictions can emerge from hierarchical representations. Furthermore, studies of human memory indicate that both priming and explicit retrieval are dependent on structured task representations.

In sum, inspection of these diverse behavioral domains consistently indicates that response selection is driven by task representations that involve more than simple S-R associations. The information guiding central processes is both conceptual and structured, as befits the intricacies of voluntary action.

11. TASK SET REPRESENTATION IN THE HUMAN BRAIN

Thus far, we have mostly discussed behavioral evidence for task set representations. Here we turn to a brief review of the neuroscience data for how task set representations may be instantiated in the brain. There are many proposed neural mechanisms for task sets. It has long been noted that the prefrontal cortex appears critical for flexible goal-based behavior (e.g., Duncan, 1986; Fuster, 1993; Lhermitte, 1983; Milner, 1963; Shallice, 1982; Stuss & Benson, 1984), so much research has focused on how this region of the brain can coordinate processes in other regions to give rise to purposeful action. We turn to data from the prefrontal cortex for clues about the task representations that might augment or supplant S-R associations in theories of voluntary behavior.

11.1 Prefrontal Cortex

Duncan (2001) noted that experiments with human participants investigating working memory and attentional control across a variety of tasks produce activity in prefrontal cortex and proposed that the response properties of neurons in prefrontal cortex (especially dorsolateral, ventrolateral, and medial regions) change according to the task demands. Some evidence for this idea comes from the ubiquity of activation in human brain regions across tasks, but more direct evidence for this adaptive coding comes from experiments on nonhuman primates by Miller and colleagues (e.g., Miller, Nieder, Freedman, & Wallis, 2003; Wallis, Anderson, & Miller, 2001). They found that some neurons along the principal sulcus in the macaque prefrontal cortex responded to task-relevant stimuli regardless of their actual identities. For example, a neuron may respond to a picture of a cat and not a dog when the cat requires a response and it may respond to the dog and not the cat when the dog requires a response. Thus, these neurons adapt to the task demands, although it is unclear whether the change in this receptivity reflects the task set, the attentional demands, or other forms of abstract coding (see also, Siegel, Buschman, & Miller, 2015). Yet, it does account for the frequencies with which prefrontal activation is observed in human studies of working memory and response selection. Some possible clues about the functional properties of the receptive fields of prefrontal neurons have been reported by Rigotti and colleagues (Barak, Rigotti, & Fusi, 2013; Rigotti et al., 2013). They proposed that the mixed selectivity of prefrontal neurons might reflect high-dimensional representations that can support flexible, complex behavior.

A large body of work has focused on the role of the prefrontal cortex in cognitive control. For example, Miller and Cohen (2001) proposed that prefrontal cortex exerts control of a situation when multiple competing representations or responses are activated. The conflict signal arises in anterior cingulate and triggers a response in prefrontal regions to adjust attention, retrieve additional information from memory or exert control in some other way to resolve the conflict. Dosenbach, Fair, Cohen, Schlaggar, and Petersen (2008) argued that control consists of multiple local modules that are connected to one another by long-range connections and organized into two main brain networks (see also, Braver, 2012). The cingulate-operculum network (including dorsal anterior cingulate and anterior and ventral prefrontal cortex) mediates task-set maintenance. This network typically mediates processing changes relatively slowly (e.g., across experimental trials). In contrast, the frontal-parietal network (including dorsolateral prefrontal and superior parietal cortices) mediates the adjustment of top-down control based on rapidly changing environmental information (e.g., within an experimental trial).

Other accounts of control involve more elaborately structured task representations. For instance, Badre (2008) proposed that the control hierarchy in prefrontal cortex is built on increasingly abstract representations of task sets. In this model, the most caudal regions (i.e., premotor cortex) represent direct S-R mappings. More anterior regions represent more abstract relationships between representations. Dorsal prefrontal cortex mediates processing when a conjunction of stimuli is required. Lateral prefrontal cortex mediates conflict/comparisons between dimensions or classes of stimuli. Anterior prefrontal cortex mediates comparisons between the context for when some set of stimulus features are task relevant or not. Koechlin and Summerfield (2007) proposed an alternative rostro-caudal model of prefrontal cortex organization (see also, Christoff & Gabrieli, 2000). Their cascade model postulates that control is implemented as a result of competition between representations at different levels of a control hierarchy. The lowest level of the hierarchy is premotor cortex. This region mediates sensory control (i.e., the S-R mappings). At the next level, dorsal prefrontal cortex mediates contextual control (i.e., selecting when particular S-R mappings should be applied over others). The next level is episodic control, which is mediated by lateral prefrontal cortex. Processing in this region overrides the current context with special case rules applicable to the current situation (episode). At the top of the control hierarchy is anterior prefrontal cortex. This region mediates branching control (i.e., maintaining not currently relevant task information for subsequent use). Finally, O'Reilly (2010) proposed that representations in prefrontal cortex change not only along the rostro-caudal dimension but also along the superior-inferior and lateral-medial dimensions. Common to these hierarchical theories is the idea that many aspects of the task and current situation are encoded (in different ways across prefrontal cortex). Although these models include S-R associations, responses are selected by a complex interplay of processes operating on multiple levels of representation that capture various components of the task.

11.2 Characterizing Control-Related Neural Activity

The complexity of these findings highlights the difficulty of interpreting neural data within the prefrontal cortex, given the many possible relationships between activation on various components of the task. In an influential review, Sakai (2008) proposed that brain regions mediating the representation of a task set should activate after a task cue and before a task stimulus (see also, Dosenbach et al., 2008). Regions with this activation profile are likely involved in the preparation for task performance (i.e., the instantiation of a task set), rather than the execution of information processing for task performance.

Dosenbach et al. (2006) sought to identify brain regions with this profile. They collected brain activation data from participants while they performed a variety of tasks. The tasks varied across stimulus modality (visual, auditory), representation category (verbal, spatial), and response modality (manual, vocal). They found that three regions (anterior prefrontal cortex, dorsal anterior cingulate cortex/medial superior frontal cortex and anterior insula/frontal operculum) that were consistently active during preparation across nearly all of their tasks. Assuming that task set representations are amodal, then these are good candidate regions to support them. However, there may be modality-specific aspects of a task set that are not captured by this amodal brain activity.

These theories of how control is instantiated in the prefrontal cortex differ in terms of the composition of the underlying processes. Some propose hierarchical influences that depend on the type of conflict within task sets; others propose distinct neural representations for slow versus fast changes in task set representation, and still others propose a general framework for how brain systems are recruited to instantiate task sets. Yet, in every case, these theories propose that the prefrontal cortex encodes complex information that is conceptual and highly structured. Complex behavior involves an interaction between multiple-levels of information—not simply associations between stimuli and responses.

12. GENERAL COMMENTS

In this chapter, we have sampled over 80 years of research showing that response selection is implemented by a complex set of processes that operate on task-based parameters as opposed to simple S-R associations. Across a range of topics, the picture is consistent: low-level stimulus and response features play a subordinate role to more abstract aspects of the task in the generation of voluntary action. Given the flexibility and complexity of human behavior, controlling motor behavior based on conceptual, structured representations seems optimal for ensuring that goals are achieved.

The emphasis on the conceptual aspects of tasks facilitates successful navigation through the real world. Goal states, in contrast to stimulus and response properties, capture the critical invariances that are necessary to encode for adaptive behavior. Humans and other animals do not rely on vagaries of the environment to present stimuli that are consistently linked to particular appropriate behaviors. Rather, we initiate motor behavior based on the end-states that we are attempting to achieve (see, Rosenbaum et al., 1990). For this reason, actions appear to be coded in terms of the expected outcomes rather than the underlying motor activations during selection (see, Hommel et al., 2001; Prinz, 1990).

The need for structure arises from the complexity of the environment and the need for the animal to produce different responses to the same stimuli depending on the current circumstances and task demands. The same movements can be made in a variety of contexts to obtain a variety of outcomes. Moreover, actions that are performed close together in time may be less related (in terms of a common goal) than actions performed far apart in time (see, e.g., Zacks & Swallow, 2007). Therefore, it is necessary to organize action to capture meaningful relationships among events that can be used to guide behavior.

With these requirements in mind, it seems reasonable to ask what is gained by referring to simple S-R associations. On a functional level, the notion of an S-R association may be inescapable. Linking particular stimuli to particular responses essentially describes most tasks. It has also been argued that S-R associations provide a key contribution to behavior, but they require additional control processes to govern their implementation (e.g., Cohen, Dunbar, & McClelland, 1990; Lhermitte, 1983; Miller & Cohen, 2001; Norman & Shallice, 1986). In this framework, abstract S-R associations that specify more than physical stimulus features and motor parameters are coupled with control processes that organize and activate them in a goal-based fashion. We are aware of no evidence that argues against this sort of approach.

Thus, we do not intend to argue that the concept necessarily be abandoned. Instead, our intention is two-fold. First, we aim to establish that theories should specify the level of representation for S-R associations, so that it is clear exactly what information forms the basis of the association. While it seems clear that in many cases the information is very abstract, the contents of such representations are most often not clearly explicated. It may be that the consideration of this issue will lead to accounts that are based on more structured associative mechanisms that are not reliant on a single level of connection mediating stimulus and response information.

Second, we pose the question of whether S-R associations are really necessary for accounts of voluntary behavior. Given the complexity of the control mechanisms already proposed, it seems possible that these processes operate on something other than S-R associations, possibly activating motor systems based on anticipated consequences. Such models would seem particularly apt for explaining novel actions, for which there should be no encoded S-R associations. If it is allowed that behavior might be driven by control processes rather than S-R associations, then practice might tune these control processes instead of instantiating S-R associations to take their place (see, Dux et al., 2009). As reviewed above, there is evidence to indicate that practice makes response selection more efficient but does not obviate its role in behavior (e.g., Dutta & Proctor, 1992; Ruthruff et al., 2003; Wifall, McMurray, & Hazeltine, 2014).

One of the obstacles to eschewing the notion of S-R associations is that it is difficult to envision alternative frameworks. Nonetheless, given the available evidence, it seems worthwhile to consider workable options. How can we integrate the importance of context and intention into models of response selection without resorting to homuncular mechanisms that choose actions based on current goal states? It may be that statistical learning processes that take into account a wide array of information (e.g., Ernst & Banks, 2002; Fiser & Aslin, 2002; Saffran, Aslin, & Newport, 1996), perhaps in conjunction with the dopaminergic reward system (e.g., Botvinick, Niv, & Barto, 2009; Miller & Cohen, 2001) and/or cognitive development (e.g., Verbruggen, McLaren, & Chambers, 2014), may be able to capture the complex patterns of behavior that we describe here. Alternatively, it might be possible to divide response selection into separate components none of which directly associate individual stimuli with responses or, at least, assign a privileged role to those associations. For example, one might conceive of two sets of processes, one based on minimizing prediction error of sensorimotor events (e.g., Rescorla & Wagner, 1972; Thompson, 1990) and another based on maximizing reward (e.g., Houk, Adams, & Barto, 1995; Izawa & Shadmehr, 2011; Schultz et al., 1995), that act in tandem to guide behavior.

In conclusion, the evidence presented here shows that S-R associations, as traditionally conceived, contribute little to theories of voluntary behavior. Motor systems rely on sensory information, and sensory systems incorporate motor states. The two sets of processes are more fundamentally integrated—with each other and with contextual information—than implied by the notion of the single bridging connection of the S-R association (see, Cisek & Kalaska, 2010; Hommel et al., 2001; Prinz, 1990). The emphasis on S-R bindings may hobble our appreciation of the richness of central processes.

ACKNOWLEDGMENTS

The authors wish to thank Matthew Bezdek, Savannah Cookson, Michael Freedberg, Christine Godwin, Susan Ravizza, and Derek Smith for their extremely helpful comments on earlier versions of this chapter.

REFERENCES

Adam, J. J., Hommel, B., & Umilta, C. (2003). Preparing for perception and action (I): the role of grouping in the response-cuing paradigm. *Cognitive Psychology, 46*, 302–358.
Akçay, Ç., & Hazeltine, E. (2007). Feature-overlap and conflict monitoring: two sources of sequential modulations. *Psychonomic Bulletin and Review, 14*, 742–748.
Akçay, Ç., & Hazeltine, E. (2008). Conflict adaptation depends on task structure. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 958–973.
Akçay, Ç., & Hazeltine, E. (2011). Domain-specific conflict adaptation without feature repetitions. *Psychonomic Bulletin and Review, 18*, 505–511.
Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: exploring the dynamic control of tasks. In C. Umilta, & M. Moscovitch (Eds.), *Attention and performance: Vol. XV. Attention and performance* (pp. 421–452). Cambridge, MA: Harvard University Press.
Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406.
Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*, 1036–1060.
Anderson, J. R., Taatgen, N. A., & Byrne, M. D. (2005). Learning to achieve perfect time-sharing: architectural implications of hazeltine, teague, and ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance, 31*, 749–761.
Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences, 12*(5), 193–200. http://dx.doi.org/10.1016/j.tics.2008.02.004.
Barak, O., Rigotti, M., & Fusi, S. (2013). The sparseness of mixed selectivity neurons controls the generalization–discrimination trade-off. *Journal of Neuroscience, 33*, 3844–3856.
Berns, G. S., Cohen, J. D., & Mintun, M. A. (1997). Brain regions responsive to novelty in the absence of awareness. *Science, 276*, 1272–1275.
Blais, C., Robidoux, S., Risko, E. F., & Besner, D. (2007). Item-specific adaptation and the conflict-monitoring hypothesis: a computational model. *Psychological Review, 114*, 1076–1086.
Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, Affective and Behavioral Neuroscience, 7*, 356–366.
Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.
Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition, 113*, 262–280.
Brass, M., & Cramon, D. Y. (2004). Decomposing components of task preparation with functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 16*, 609–620.
Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences, 16*, 106–113.
Buss, A., Wifall, T., Hazeltine, E., & Spencer, J. P. (2014). Integrating the behavioral and neural dynamics of response selection in a dual-task paradigm: a dynamic neural field model of Dux et al. (2009). *Journal of Cognitive Neuroscience, 26*, 334–351.
Christoff, K., & Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology, 28*(2), 168–186.
Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron, 45*, 801–814.
Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interaction with a world full of action choices. *Annual Review of Neuroscience, 33*, 269–298.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332–361.
Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information-processing. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 529–553.
De Jong, R., Liang, C.-C., & Lauber, E. J. (1994). Conditional and unconditional automaticity: a dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 731–750.
Diedrichsen, J., Hazeltine, E., Kennerley, S., & Ivry, R. B. (2001). Absence of bimanual interference during directly-cued actions. *Psychological Science, 12*, 493–498.
Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences, 12*(3), 99–105. http://dx.doi.org/10.1016/j.tics.2008.01.001.
Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. S. C., et al. (2006). A core system for the implementation of task sets. *Neuron, 50*(5), 799–812. http://dx.doi.org/10.1016/j.neuron.2006.04.031.
Doyon, J., Gaudreau, D., Laforce, R. J., Catronguay, M., Bedard, F., & Bouchard, J.-P. (1997). Role of striatum, cerebellum, and frontal lobes in the learning of a visuomotor sequence. *Brain and Cognition, 34*, 218–245.
Dreisbach, G., Goschke, T., & Haider, H. (2006). Implicit task sets in task switching. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32*, 1221–1233.
Dreisbach, G., Goschke, T., & Haider, H. (2007). The role of task rules and stimulus-response mappings in the task switching paradigm. *Psychological Research, 71*, 383–392.
Dreisbach, G., & Haider, H. (2008). That's what task sets are for: shielding against irrelevant information. *Psychological Research, 72*, 355–361.
Dreisbach, G., & Haider, H. (2009). How task representations guide attention: further evidence for the shielding function of task sets. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*, 477–486.
Duncan, J. (1977a). Response selection errors in spatial choice reaction tasks. *Quarterly Journal of Experimental Psychology, 29*, 415–423.
Duncan, J. (1977b). Response selection rules in spatial choice reaction tasks. In S. Dornic (Ed.), *Attention and performance* (Vol. VI, pp. 49–61). Hillsdale: NJ: Erlbaum.
Duncan, J. (1978). Response selection in spatial choice reaction: further evidence against associative models. *The Quarterly Journal of Experimental Psychology, 30*, 429–440.
Duncan, J. (1986). Disorganisation of behaviour after frontal lobe damage. *Cognitive Neuorpsychology, 3*, 271–290.
Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2*(11), 820–829. http://dx.doi.org/10.1038/35097575.
Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neuroscience, 23*(10), 475–483.
Dutta, A., & Proctor, R. W. (1992). Persistence of stimulus-response compatibility effects with extended practice. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 801–809.
Dux, P. E., Ivanoff, J., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron, 52*, 1109–1120.
Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron, 63*, 127–138.
Egner, T., Delano, M., & Hirsch, J. (2007). Separate conflict-specific cognitive control mechanisms in the human brain. *NeuroImage, 35*, 940–948.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience, 8*, 1784–1790.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 229–240.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*, 143–149.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: a continuous flow conception and experimental results. *Perception and Psychophysics, 25*, 249–263.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Fernandez-Duque, D., & Knight, M. (2008). Cognitive control: dynamic, sustained, and voluntary influences. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 340–355.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 458–467.

Fitts, P. M., & Deininger, R. L. (1954). S-R Compatibility: correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology, 48*, 483–492.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46*, 199–210.

Franz, E. A., Eliassen, J. C., Ivry, R. B., & Gazzaniga, M. S. (1996). Dissociation of spatial and temporal coupling in the bimanual movements of callosotomy patients. *Psychological Science, 7*, 306–310.

Freedberg, M. V., Wagschal, T. T., & Hazeltine, E. (2014). Incidental learning and task boundaries. *Journal of Experimental Psychology: Learning, Memory and Cognition, 40*, 1680–1700.

Freitas, A. L., Bahar, M., Yang, S., & Banai, R. (2007). Contextual adjustments in cognitive control across tasks. *Psychological Science, 18*, 1040–1043.

Frensch, P. A., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous serial transactions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 567–584.

Funes, M. J., Lupiáñez, J., & Humphreys, G. W. (2010). Analyzing the generality of conflict adaptation effects. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 147–161.

Fuster, J. M. (1993). Frontal lobes. *Current Opinion in Neurobiology, 3*, 160–165.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (2002). Action representation and the inferior parietal lobule. In W. Prinz, & B. Hommel (Eds.), *Attention and performance* (Vol. XIX). Oxford: Oxford University Press.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science, 4*, 123–124.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology, 45*, 220–266.

Goodman, D., & Kelso, J. A. (1980). Are movements prepared in parts? Not under compatible (naturalized) conditions. *Journal of Experimental Psychology: General, 109*, 475–495.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience, 7*, 497–510.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1998). Abstract and effector-specific representations of motor sequences identified with PET. *Journal of Neurophysiology, 18*, 9420–9428.

Gratton, G., Coles, M. G. H., & Donchin, O. (1992). Optimizing the use of information: strategic control of activation and responses. *Journal of Experimental Psychology: General, 121*, 480–506.

Greenwald, A. G. (1970). A choice reaction time test of ideomotor theory. *Journal of Experimental Psychology, 86*, 20–25.
Greenwald, A. G. (1972). On doing two things at once: time sharing as a function of ideomotor compatibility. *Journal of Experimental Psychology, 94*, 52–57.
Guiard, Y. (1983). The lateral coding of rotations: a study of the Simon effect with wheel-rotation responses. *Journal of Motor Behavior, 15*, 331–342.
Halvorson, K. M., Ebner, H., & Hazeltine, E. (2013). Investigating perfect time-sharing: the relationship between IM-compatible tasks and dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 527–545.
Halvorson, K. M., & Hazeltine, E. (2015). Do small dual-task costs reflect ideomotor compatibility or the absence of crosstalk? *Psychonomic Bulletin and Review*. http://dx.doi.org/10.3758/s13423-015-0813-8.
Halvorson, K. M., Wagschal, T. T., & Hazeltine, E. (2013). Conceptualization of task boundaries preserves implicit sequence learning under dual-task conditions. *Psychonomic Bulletin and Review, 20*, 1005–1010.
Hammond, N., & Barber, P. J. (1978). Evidence for abstract response codes: ear-hand correspondence effects in a three choice reaction-time task. *Quarterly Journal of Experimental Psychology, 30*, 71–82.
Hawkins, H. L., MacKay, S. L., Holley, S. L., Friedin, B. D., & Cohen, S. L. (1973). Locus of the relative frequency effect in choice reaction time. *Journal of Experimental Psychology, 101*, 90–99.
Hazeltine, E. (2002). The representational nature of sequence learning: evidence for goal-based codes. In W. Prinz, & B. Hommel (Eds.), *Attention and performance* (Vol. XIX, pp. 673–689). Oxford: University Press.
Hazeltine, E. (2005). Response-response compatibility during bimanual movements: evidence for the conceptual coding of action. *Psychonomic Bulletin and Review, 12*, 682–688.
Hazeltine, E., Akçay, Ç., & Mordkoff, J. T. (2011a). Keeping Simon simple: examining the relationship between sequential modulations and feature repetitions with two stimuli, two locations, and two responses. *Acta Psychologia, 136*, 245–252.
Hazeltine, E., Diedrichsen, J., Kennerley, S., & Ivry, R. B. (2003). Bimanual cross-talk during reaching movements is primarily related to response selection, not the specification of motor parameters. *Psychological Research, 67*, 56–70.
Hazeltine, E., Grafton, S. T., & Ivry, R. B. (1997). Attention and stimulus characteristics determine the locus of motor sequence encoding: a PET study. *Brain, 120*, 123–140.
Hazeltine, E., & Ivry, R. B. (2002). Neural structures that support implicit sequence learning. In L. Jimenez (Ed.), *Attention and implicit learning* (pp. 71–107). Amsterdam: John Benjamins.
Hazeltine, E., Lightman, E., Schwarb, H., & Schumacher, E. H. (2011b). The boundaries of sequential modulations: evidence for set-level control. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1898–1914.
Hazeltine, E., & Ruthruff, E. (2006). Modality pairing effects and the response selection bottleneck. *Psychological Research, 70*, 504–513.
Hazeltine, E., Ruthruff, E., & Remington, R. W. (2006). The role of input and output modality pairings in dual-task performance: evidence for content-dependent central interference. *Cognitive Psychology, 52*, 291–345.
Hazeltine, E., Teague, D., & Ivry, R. B. (2002). Simultaneous dual-task performance reveals parallel response selection after practice. *Journal of Experimental Psychology: Human Perception and Performance, 28*(3), 527–545.
Hazeltine, E., & Wifall, T. (2011). Searching working memory for the source of dual-task costs. *Psychological Research, 75*, 466–475.
Hedge, A., & Marsh, N. W. A. (1975). The effect of irrelevant spatial correspondence on two-choice response time. *Acta Psychologia, 39*, 340–347.

Henson, R. N., Eckstein, D., Waszak, F., Frings, C., & Horner, A. (2014). Stimulus–response bindings in priming. *Trends in Cognitive Science, 18*, 376–384.
Herath, P., Klingberg, T., Young, J., Amunts, K., & Roland, P. (2001). Neural correlates of dual task interference can be dissociated from those of divided attention: an fMRI study. *Cerebral Cortex, 11*, 796–805.
Heuer, H. (1995). Models for response-response compatibility: the effects of the relation between responses in a choice task. *Acta Psychologia, 90*, 315–332.
Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11–26.
Holland, P. C. (1983). "Occasion-setting" in Pavlovian feature positive discriminations. In R. J. Hernstein, & A. R. Wagner (Eds.), *Discrimination processes: Vol. 4. Quantitative analyses of behavior* (pp. 183–206). Cambridge, MA: Ballinger.
Hommel, B. (1993). Inverting the Simon effect by intention: determinants of direction and extent effects of irrelevant spatial information. *Psychological Research, 55*, 270–279.
Hommel, B. (1998). Automatic stimulus-response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1368–1384.
Hommel, B. (2000). The prepared reflex: automaticity and control in stimulus-response translation (tutorial). In S. Monsell, & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 247–274). Cambridge, MA: MIT Press.
Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends in Cognitive Science, 8*, 494–500.
Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologia, 136*, 189–202.
Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC). *Behavioral and Brain Sciences, 24*, 849–878.
Horner, A., & Henson, R. N. (2009). Bindings between stimuli and multiple response codes dominate long-lag repetition priming in speeded classification tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*, 757–779.
Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge, MA: MIT Press.
Huestegge, L., & Koch, I. (2010). Crossmodal action selection: evidence from dual-task compatibility. *Memory and Cognition, 38*, 493–501.
Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*, 188–196.
Ivry, R. B., Franz, E. A., Kingstone, A., & Johnston, J. C. (1998). The psychological refractory period effect following callosotomy: uncoupling of lateralized response codes. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 463–480.
Ivry, R. B., & Hazeltine, E. (2000). Task switching in a callosotomy patient and normal participants: evidence for response-related sources of interference. In S. Monsell, & J. Driver (Eds.), *Attention and performance* (Vol. XVIII, pp. 401–423). Cambridge, MA: MIT Press.
Izawa, J., & Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Computational Biology,* 7, e1002012. http://dx.doi.org/10.1371/journal.pcbi.1002012.
Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology, 89*.
Jiang, Y., & Kanwisher, N. (2003a). Common neural mechanisms for response selection and perceptual processing. *Journal of Cognitive Neuroscience, 15*, 1095–1110.
Jiang, Y., & Kanwisher, N. (2003b). Common neural substrates for response selection across modalities and mapping paradigms. *Journal of Cognitive Neuroscience, 15*, 1080–1094.
Jiang, Y., Saxe, R., & Kanwisher, N. (2004). Functional magnetic resonance imaging provides new constraints on theories of the psychological refractory period. *Psychological Science, 15*, 390–396.

Kamin, L. J. (1968). Attention-like processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation*. Miami: University of Miami Press.
Keele, S. W. (1970). Effects of input and output modes on decision time. *Journal of Experimental Psychology, 85*, 157–164.
Keele, S. W., Ivry, R. B., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110*, 316–339.
Keele, S. W., Jennings, P., Jones, S., Caulton, D., & Cohen, A. (1995). On the modularity of sequence representation. *Journal of Motor Behavior, 27*, 17–30.
Kelso, J. A., Southard, D. L., & Goodman, D. (1979). On the coordination of two-handed movements. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 229–238.
Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*, 153–161.
Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., et al. (2010). Control and interference in task switching—a review. *Psychological Bulletin, 136*, 849–874.
Kiesel, A., Wendt, M., & Peters, A. (2007). Task switching: on the origin of response congruency effects. *Psychological Research, 71*, 117–125.
Kintsch, W. (1970). *Learning, memory, and conceptual processes*. New York: Wiley.
Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. *Journal of Memory and Language, 28*, 56–77.
Kleinsorge, T., & Heuer, H. (1999). Hierarchical switching in a multi-dimensional task space. *Psychological Research, 62*, 300–312.
Klingberg, T. (1998). Concurrent performance of two working memory tasks: potential mechanisms of interference. *Cerebral Cortex, 8*, 593–601.
Koch, I., & Allport, A. (2006). Cue-based preparation and stimulus-based priming of tasks in task switching. *Memory and Cognition, 34*, 433–444.
Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences, 11*(6), 229–235. http://dx.doi.org/10.1016/j.tics.2007.04.005.
Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility–A model and taxonomy. *Psychological Review, 97*, 253–270.
Kornblum, S., & Lee, J.-W. (1995). Stimulus-response compatibility with relevant and irrelevant stimulus dimensions that do and do not overlap with the response. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 855–875.
Laberge, D., Legrand, R., & Hobbie, R. K. (1969). Functional identification of perceptual and response biases in choice reaction time. *Journal of Experimental Psychology, 79*, 295–299.
Laberge, D., & Tweedy, J. R. (1964). Presentation probability and choice time. *Journal of Experimental Psychology, 68*, 477–481.
Lhermitte, F. (1983). 'Utilization behaviour' and its relation to lesions of the frontal lobes. *Brain, 106*, 237–255.
Liefooghe, B., Wenke, D., & De Houwer, J. (2012). Instruction-based task-rule congruency effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*, 1325–1335.
Lien, M.-C., & Proctor, R. W. (2002). Stimulus-response compatibility and psychological refractory period effects: implications for response selection. *Psychonomic Bulletin and Review, 9*(2), 212–238.
Lien, M.-C., & Ruthruff, E. (2004). Task switching in a hierarchical task structure: evidence for the fragility of the task repetition benefit. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 697–713.

Lien, M.-C., Ruthruff, E., Remington, R. W., & Johnston, J. C. (2005). On the limits of advance preparation for a task-switch: do people prepare all of the task some of the time or some of the task all the time. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 299–315.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Logan, G. D. (1990). Repetition priming and automaticity: common underlying mechanisms? *Cognitive Psychology, 22*, 1–35.

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance, 29*, 575–599.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108*, 393–434.

Mayr, U., & Bryck, R. L. (2005). Sticky rules: integration between abstract rules and specific actions. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 337–350.

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 362–372.

McCann, R. S., & Johnston, J. C. (1992). Locus of the single-channel bottleneck in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 471–484.

McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review, 31*, 133–156.

Mechsner, F., Kerzel, D., Knoblich, G., & Prinz, W. (2001). Perceptual basis of bimanual coordination. *Nature, 414*, 69–73.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory phenomena. *Psychological Review, 104*, 749–791.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of human multiple task performance: the EPIC information-processing architecture and strategic response deferment model. *Psychological Review, 104*, 1–65.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Miller, J. (1982). Discrete versus continuous models of human information processing: in search of partial output. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 273–296.

Miller, J. (1985). A hand advantage in preparation of simple keypress responses: reply to Reeve and Proctor (1984). *Journal of Experimental Psychology: Human Perception and Performance, 11*, 221–233.

Miller, J. (1988). Discrete and continuous models of human information processing: theoretical distinctions and empirical results. *Acta Psychologica, 67*, 191–257.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.

Miller, E. K., Nieder, A., Freedman, D. J., & Wallis, J. D. (2003). Neural correlates of categories and concepts. *Current Opinion in Neurobiology, 13*, 198–203.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology, 9*, 90–100.

Monsell, S., & Driver, J. (2000). Banishing the control homunculus. In S. Monsell, & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 3–32). Cambridge, Massachusetts: MIT Press.

Mordkoff, J. T., & Hazeltine, E. (2011). Parallel patterns of spatial compatibility and spatial congruence..as long as you don't look too closely. *Acta Psychologia, 136*, 253–258.

Müsseler, J., & Hommel, B. (1997). Blindness to response-compatible stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 861–872.

Nagel, I. E., Schumacher, E. H., Goebel, R., & D'Esposito, M. (2008). Functional MRI investigation of verbal selection mechanisms in lateral prefrontal cortex. *NeuroImage, 43*, 801–807.

Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 435–448.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology, 19*, 1–32.

Norman, D. A., & Shallice, T. (1986). Attention to action: willed and automatic control of behavior. In J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1–18). New York: Plenum.

Notebaert, W., & Verguts, T. (2008). Cognitive control acts locally. *Cognition, 106*, 1071–1080.

O'Reilly, R. C. (2010). The what and how of prefrontal cortical organization. *Trends in Neurosciences, 33*, 355–361.

Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 358–377.

Pashler, H. (1989). Dissociations and dependencies between speed and accuracy: evidence for a two-component theory of divided attention in simple tasks. *Cognitive Psychology, 21*, 469–514.

Pashler, H. (1990). Do response modality effects support multiprocessor models of divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 826–842.

Pashler, H. (1994a). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin, 116*, 220–244.

Pashler, H. (1994b). Graded capacity-sharing in dual-task interference? *Journal of Experimental Psychology: Human Perception and Performance, 20*, 330–342.

Pashler, H., & Baylis, G. (1991). Procedural learning: 1. Locus of practice effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*(1), 20–32.

Pashler, H., & Johnston, J. C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *Quarterly Journal of Experimental Psychology, 41A*, 19–45.

Pinel, J. P. J., & Treit, D. (1979). Conditioned defensive burying in rats: availability of burying materials. *Animal Learning and Behavior, 7*, 392–396.

Pollack, I. (1959). Message uncertainty and message reception. *The Journal of the Acoustical Society of America, 31*, 1500–1508.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann, & W. Prinz (Eds.), *Relationships between perception and action* (pp. 167–201). Berlin: Springer-Verlag.

Proffitt, D. R. (2006). Distance perception. *Current Directions in Psychological Science, 15*, 131–135.

Rabbitt, P. M. A. (1959). Effects of independent variations in stimulus and response probability. *Nature, 183*, 1212.

Rabbitt, P. M. A., & Vyas, S. M. (1973). What is repeated in the "repetition effect"?. In S. Kornblum (Ed.), *Attention and peformance* (Vol. IV, pp. 327–342). London: Academic Press.

Rauch, S. L., Savage, C. R., Brown, H. D., Curran, T., Alpert, N. M., Kendrick, A., & Kosslyn, S. M. (1995). A PET investigation of implicit and explicit sequence learning. *Human Brain Mapping, 3*, 271–286.

Reeve, T. G., & Proctor, R. W. (1984). On the advance preparation of discrete finger responses. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 541–553.

Reeve, T. G., & Proctor, R. W. (1985). Non-motoric translation processes in the preparation of discrete finger responses: a rebuttal of Miller's (1985) analysis. *Journal of Expermental Psychology: Human Perception and Performance, 11*, 234–240.

Reeve, T. G., & Proctor, R. W. (1990). The salient-features coding principle for spatial- and symbolic compatibility effects. In R. W. Proctor, & T. G. Reeve (Eds.), *Stimulus-response compatibility* (pp. 163–180). Amsterdam: North-Holland.

Reisberg, D., Baron, J., & Kemler, D. G. (1980). Overcoming Stroop interference: the effects of practice on distractor potency. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 140–150.

Rescorla, R. A. (1972). Second-order conditioning: implications for theories of learning. In F. J. McGuigan, & D. Lumsden (Eds.), *Contemporary approaches to conditioning and learning*. New York: Winston.

Rescorla, R. A. (1988a). Behavioral studies of Pavlovian conditioning. *Annual Review of Neuroscience, 11*, 329–352.

Rescorla, R. A. (1988b). Pavlovian conditioning: it's not what you think it is. *American Psychologist, 43*, 151–160.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., et al. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature, 497*, 585–590.

Rogers, R., & Monsell, S. (1995). The costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: Human Perception and Performance, 124*, 207–231.

Rosenbaum, D. A. (1980). Human movement initiation: specification of arm, direction, and extent. *Journal of Experimental Psychology: General, 109*, 444–474.

Rosenbaum, D. A. (1983). The movement precuing technique: assumptions, applications, and extensions. In R. A. Magill (Ed.), *Memory and control of action* (pp. 230–274). Amsterdam: North-Holland Publishing Company.

Rosenbaum, D. A., Marchak, F., Barnes, H. J., Vaughan, J., Slotta, J., & Jorgensen, M. (1990). Constraints for action selection: overhand versus underhand grips. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 321–342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience, 12*(4), 664–678. http://dx.doi.org/10.1162/089892900562291.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4*(3), 108–115. http://dx.doi.org/10.1016/s1364-6613(00)01445-5.

Ruthruff, E., Johnston, J. C., & Van Selst, M. (2001). Why practice reduces dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 3–21.

Ruthruff, E., Johnston, J. C., Van Selst, M., Whitsell, S., & Remington, R. W. (2003). Vanishing dual-task interference after practice: Has the bottleneck been eliminated or is it merely latent? *Journal of Experimental Psychology: Human Perception and Performance, 29*, 280–289.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Sakai, K. (2008). Task set and prefrontal cortex. *Annual Review of Neuroscience, 31*, 219–245. Palo Alto: Annual Reviews.

Salvucci, D. D., & Taatgen, N. A. (2010). *The multitasking mind*. New York: Oxford University Press.

Schmidt, J. R., & Weissman, D. H. (2014). Congruency sequence effects without feature integration or contingency learning confounds. *PLoS One, 9*.
Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M. I. Posner, & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 475–492). Hillsdale, NJ: Erlbaum.
Schneider, D. W. (2014). Isolating a mediated route for response congruency effects in task switching. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*, 235–245.
Schneider, D. W., & Anderson, J. R. (2011). A memory-based mode of Hick's law. *Cognitive Psychology, 62*, 193–222.
Schneider, D. W., & Logan, G. D. (2007). Retrieving information from a hierarchical plan. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*, 1076–1091.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.
Schultz, W., Romo, R., Ljungberg, T., Mirenowicz, J., Hollerman, J. R., & Dickinson, A. (1995). Reward-related signals carried by dopamine neurons. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 233–248). Cambridge, MA: MIT Press.
Schumacher, E. H., & D'Esposito, M. (2002). Neural implementation of response selection in humans as revealed by localized effects of stimulus-response compatibility on brain activation. *Human Brain Mapping, 17*, 193–201.
Schumacher, E. H., Elston, P. A., & D'Esposito, M. (2003). Neural evidence for representation-specific response selection. *Journal of Cognitive Neuroscience, 15*, 1111–1121.
Schumacher, E. H., & Schwarb, H. (2009). Parallel response selection disrupts sequence learning under dual-task conditions. *Journal of Experimental Psychology: General, 138*, 270–290.
Schumacher, E. H., Schwarb, H., Lightman, E., & Hazeltine, E. (2011). Investigating the modality specificity of response selection using a temporal flanker task. *Psychological Research, 75*, 499–512.
Schumacher, E. H., Seymour, T. L., Glass, J. M., Kieras, D. E., & Meyer, D. E. (2001). Virtually perfect time sharing in dual-task performance: uncorking the central attentional bottleneck. *Psychological Science, 12*, 101–108.
Schwarb, H., & Schumacher, E. H. (2009). Neural evidence of a role for spatial response selection in the learning of spatial sequences. *Brain Research, 1247*, 114–125.
Schwarb, H., & Schumacher, E. H. (2010). Implicit sequence learning is represented by stimulus-response rules. *Memory and Cognition, 38*, 677–688.
Schwarb, H., & Schumacher, E. H. (2012). Generalized lessons about sequence learning from the study of the serial reaction time task. *Advances in Cognitive Psychology, 8*, 165–178.
Seidler, R. D., Purushotham, A., Kim, S.-G., Ugurbil, K., Willingham, D. B., & Ashe, J. (2002). Cerebellum activation associated with performance change but not motor learning. *Science, 296*, 2043–2046.
Shaffer, L. H. (1965). Choice reaction with variable S-R mapping. *Journal of Experimental Psychology, 70*, 284–288.
Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B, 298*, 199–209.
Siegel, M., Buschman, T. J., & Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science, 384*, 1352–1355.
Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology, 81*, 174–176.
Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: the effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*, 300–304.
Smith, M. C. (1967). Theories of the psychological refractory period. *Psychological Bulletin, 67*, 202–213.

Soto, F. A., & Wasserman, E. A. (2012). A category-overshadowing effect in pigeons: support for the common elements model of object categorization learning. *Journal of Experimental Psychology: Animal Behavior Processes, 38*, 322–328.
Spapé, M. M., & Hommel, B. (2008). He said, she said: episodic retrieval induces conflict adaptation in an auditory Stroop task. *Psychonomic Bulletin and Review, 15*, 1117–1121.
Spivey, M. J., Dale, R., Knoblich, G., & Grosjean, M. (2010). Do curved reaching movements emerge from competing perceptions? *Journal of Expermental Psychology: Human Perception and Performance, 36*, 251–254.
Stelzel, C., Schumacher, E. H., Schubert, T., & D'Esposito, M. (2006). The neural effect of stimulus-response modality compatibility on dual-task performance: an fMRI study. *Psychological Research, 70*, 514–525.
Sternberg, S. (1969). The discovery of processing stages: extension of Donder's method. In W. G. Koster (Ed.), *Attention and performance* (Vol. II, pp. 276–315). Amsterdam: North Holland.
Stoffels, E. J. (1996). Uncertainty and processing routes in the selection of a response. An S-R compatibility study. *Acta Psychologia, 94*, 227–252.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95*, 3–28.
Swinnen, S. (2002). Intermanual coordination: from behavioural principles to neural-network interactions. *Nature Reviews Neuroscience, 3*, 350–361.
Szameitat, A. J., Schubert, T., Müller, K., & von Cramon, D. Y. (2002). Localization of executive function in dual-task performance with fMRI. *Journal of Cognitive Neuroscience, 14*, 1184–1199.
Taylor, J. A., & Ivry, R. B. (2013). Context-dependent generalization. *Frontiers in Human Neuroscience,* 7(171). http://dx.doi.org/10.3389/fnhum.2013.00171.
Theios, J. (1975). The components of response latency in simple human information processing tasks. In P. M. A. Rabbitt, & S. Dornic (Eds.), *Attention and performance* (Vol. V, pp. 418–440). London: Academic Press.
Thompson, R. F. (1990). Neural mechanisms of classical conditioning in mammals. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 329*, 161–170.
Toni, I., Krams, M., Turner, R., & Passingham, R. E. (1998). The time course of changes during motor sequence learning: a whole-brain fMRI study. *NeuroImage, 8*, 50–61.
Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict adaptation effect: it's not just priming. *Cognitive, Affective and Behavioral Neuroscience, 5*, 467–472.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review, 108*, 550–592.
Verbruggen, F., McLaren, I. P. L., & Chambers, C. D. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspectives in Psychological Science, 9*, 497–524.
Verguts, T., & Notebaert, W. (2008). Hebbian learning and cognitive control: dealing with specific and non-specific adaptation. *Psychological Review, 115*, 518–525.
Wallace, R. J. (1971). S-R compatibility and the idea of a response code. *Journal of Experimental Psychology, 88*, 354–360.
Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in the prefrontal cortex encode abstract rules. *Nature, 411*, 953–956.
Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology, 46*, 361–413.
Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104*, 442–452.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance—a review and a theory. *British Journal of Psychology, 434*, 2—19.
Welford, A. T. (1958). *Ageing and human skill*. London: Oxford University Press.
Welford, A. T. (1967). Single channel operation in the brain. *Acta Psychologia, 27*, 5—22.
Wenke, D., Gaschler, R., & Nattkemper, D. (2007). Instruction-induced feature binding. *Psychological Research, 71*, 92—106.
Wickens, D. D. (1970). Encoding categories of words: an empirical approach to meaning. *Psychological Review, 77*, 1—15.
Wickens, D. D. (1972). Characteristics of word encoding. In A. W. Melton, & E. Martin (Eds.), *Coding processes in human memory* (pp. 191—215). Washington, D.C.: Winston/Wiley.
Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 2*(5—6), 440—445.
Wifall, T., Hazeltine, E., & Mordkoff, J. T. (2015). The roles of stimulus and response uncertainty in forced-choice performance: an amendment of Hick/Hyman Law. *Psychological Research*. http://dx.doi.org/10.1007/s00426-015-0675-8.
Wifall, T., McMurray, B., & Hazeltine, E. (2014). Perceptual similarity affects the learning curve (but not necessarily learning). *Journal of Experimental Psychology: General, 143*, 312—331.
Wilding, E. L. (1999). Separating retrieval strategies from retrieval success: an event-related potential study of source memory. *Neuropsychologia, 37*(4), 441—454. http://dx.doi.org/10.1016/s0028-3932(98)00100-6.
Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review, 105*, 558—584.
Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 1047—1060.
Willingham, D. B., Wells, L. A., Farrell, J. M., & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory and Cognition, 28*, 366—375.
Wright, C. E. (1990). Generalized motor programs: reevaluating claims of effector independence. In M. Jeannerod (Ed.), *Attention and performance* (Vol. XIII, pp. 294—320). Hillsdale, NJ: Lawrence Erlbaum.
Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*, 80—84.
Ziessler, M., & Nattkemper, D. (2001). Learning of event sequences is based on response effect learning: further evidence from serial reaction task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 595—613.
Ziessler, M., & Nattkemper, D. (2002). Effect anticipation in action planning. In W. Prinz, & B. Hommel (Eds.), *Attention and performance* (Vol. XIX, pp. 645—672). Oxford: University Press.

This page intentionally left blank

CHAPTER SEVEN

What Dot-Based Masking Effects Can Tell Us About Visual Cognition: A Selective Review of Masking Effects at the Whole-Object and Edge-Based Levels

Todd A. Kahan[1]
Department of Psychology, Bates College, Lewiston, ME, USA
[1]Corresponding author: E-mail: tkahan@bates.edu

Contents

Abstract

Vision is a multistep process that begins when light stimulates cells in the retina, with multiple processes quickly cascading throughout the brain, providing new information to inform potential actions. Complicating this is the fact that vision does not only proceed unidirectionally, or bottom-up, but in every region of the brain, reentrant or

Psychology of Learning and Motivation, Volume 64
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2015.09.007

top-down signals influence earlier stages by feeding activity back to previously activated brain regions. This reentrant processing is both iterative and dynamic. In this chapter I review several new developments in dot-based masking that serve as tools for studying how this dynamic flow of information helps us to make sense of the visual environment. Four-dot object-substitution masking, which arises at the whole-object level, is contrasted with two newly discovered edge-based masking effects (object trimming and object binding). I then review the evidence which shows that by examining these effects, researchers are gaining a more solid understanding of the role reentrant processes play in conscious awareness, the role reentrant processes play in feature binding, the level at which objects are processed when the reentrant sweep begins, and whether reentrant pathways carry top-down information about meaning and cultural expectations that can guide the edge-assembly process before objects have even been recognized.

1. INTRODUCTION

To many psychologists visual masking effects may, at first glance, seem rather mundane. In fact, it is not at all surprising that visual perception of a target is impaired by a temporally and spatially nearby mask. Indeed, if this were the end of the story then the relative short-shrift visual masking effects are given in most perception textbooks, where masking is quite often treated as nothing more than a curiosity of the visual system, would seem justified. But this could not be further from the truth. In this chapter I show how several recent developments in the masking literature have helped to improve our understanding of vision and cognition generally.

In 1997, Enns and Di Lollo identified a new form of visual masking which has, over the years, been referred to as common-onset masking, four-dot masking, and/or object-substitution masking (OSM). Since that time the terms four-dot masking and OSM have been used most frequently. Here I refer to this effect as four-dot OSM. Similarities and differences between four-dot OSM and other types of masking have been reviewed elsewhere (see Enns, 2004; Enns & Di Lollo, 2000) as have the data which help to delineate the cause of this type of masking (see Enns, 2008; Goodhew, Pratt, Dux, & Ferber, 2013). As such, the current chapter does not review these issues in any depth. Instead, this chapter has two goals: (1) to outline several new developments in dot-based masking effects, namely the discovery of two edge-based effects which my colleagues and I refer to as object trimming and object binding and (2) to show how dot-based masking at the whole-object and edge-based levels can be used as tools for gaining a better understanding of visual cognition.

In what follows I first review several key findings related to dot-based masking by distinguishing between four-dot OSM, which arises at the whole-object level, and two-dot masking effects, which cause perceptual distortions at the edge-based level (i.e., object trimming and object binding). Following this I explore evidence which firmly demonstrates that by leveraging what is known about these different dot-based masking effects researchers can now, for the first time, answer large-scale questions related to object recognition, including, but not limited to: the role of reentrant processes in awareness and feature binding, the role of reentrant processes in the extraction of meaning from stimuli, the top-down influence of meaning on early stages of visual perception, and the role of cultural expectations on visual perception. As I review this evidence I highlight areas that need further research and I conclude with a listing of what has been learned thus far. It is an exciting time to be a vision scientist. It is my hope that this chapter will make it clear how dot-based masking effects can provide useful insights into how our visual system uses feedforward and reentrant pathways in a dynamic and ongoing manner to help us to make sense of the visual world.

2. VISION AND DOT-BASED MASKING

When light from an object enters the eye, a chain reaction of neural activity is started that begins with the feedforward extraction of an object's basic-level features (edges, color, etc.) and is followed by the assembly of these features into a complete object form. But the exchange of information within the visual system is not only unidirectional. Indeed, even before the neural registration of new information in any given region of the brain, that region is under the influence of past events. Moreover, the rapid feedforward sweep of new neural activity is very quickly accompanied by a resurgence of top-down activity, and this bidirectional and iterative exchange of information along bottom-up and top-down pathways is what, under normal viewing conditions facilitates object recognition, but under lab-based conditions gives rise to four-dot OSM.

2.1 Four-Dot OSM and Reentry at the Whole-Object Level

Four-dot OSM is the robust finding that a briefly shown target object that is surrounded by four small dots is much more difficult to identify when the dots remain visible after target offset relative to when the dots and target appear and disappear together (see Figure 1). The dominant explanation as to why this effect occurs is based on reentrant pathways in the visual system (Di Lollo, Enns, & Rensink, 2000, 2002). According to this explanation, the target and mask are extracted and then this information is

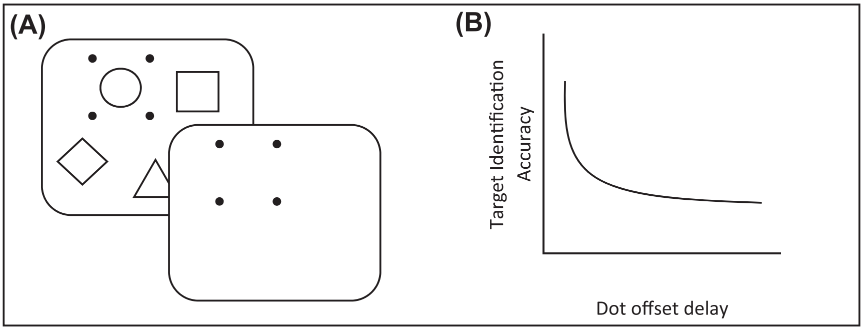

Figure 1 When a target shape (here a circle) is surrounded by four small dots (Panel A), accuracy at identifying this target decreases as a function of dot offset delay (Panel B). The hypothetical data shown in Panel B depict the typical pattern of results[1].

entered into a developing perceptual hypothesis analogous to an "object file" (Kahneman & Treisman, 1984). This developing percept is then updated and checked, in an iterative and ongoing manner, against incoming visual information by way of reentrant pathways which descend to lower-level brain areas (Mumford, 1991, 1992). The purpose of this cyclical updating process is to boost signal strength and reduce noise, and when the target and dots appear and disappear together accuracy is quite high. However, when the dots remain visible following target offset, the iterative exchange of information results in the target-plus-mask representation, which was originally extracted during the initial feedforward pass, being replaced with the mask alone, and accuracy is poor (see Panel B in Figure 1)[2]. The physiological evidence for reentrant signals in the brain is clear (Bullier, 2001a; Bullier, McCourt, & Henry, 1988; Felleman & Van Essen, 1991; Mignard & Malpeli, 1991) and to date, there is now widespread behavioral, electrophysiological, brain imaging, and computational modeling evidence supporting the role of reentrant processing in four-dot OSM (for reviews, see Enns, Lleras, & Di Lollo, 2006; Di Lollo,

[1] Although performance has been shown to generally decrease as mask duration increases, and this is the pattern I depict here, Goodhew et al. (2011) report a slight increase in target accuracy at very long mask durations (not shown in this figure). These authors make the argument that at very long mask durations lingering target activation can be used to instantiate the target as a separate and distinct object from the mask.

[2] Purely feedforward explanations have also been proposed to explain four-dot OSM (see Francis & Hermens, 2002 as well as Põder, 2013) but because these accounts cannot explain the entirety of the available evidence (see Di Lollo et al., 2002 and Di Lollo, 2014), reentrant accounts of four-dot OSM are now much more widely accepted.

2014). Most notably, when repetitive transcranial magnetic stimulation (rTMS) has been used to disrupt reentrant communication between brain areas target accuracy in the delayed-offset condition is selectively improved and four-dot OSM is reduced (i.e., the difference in performance between immediate and delayed offset is lessened) relative to performance pre-rTMS or with sham stimulation (Hirose et al., 2005, 2007). These results add to the large body of experimental evidence that now supports four-dot OSM being caused by reentry.

That four-dot OSM is caused by a mask that has a delayed offset relative to the target and involves reentrant processing does not, by itself, indicate that four-dot OSM reflects the updating of an object-level representation (i.e., that this effect arises at the whole-object level), rather than the updating of many disjointed features that have not yet been combined to form a cohesive whole. Nevertheless, there exist several behavioral and physiological experiments that speak to this issue.

Some of the more convincing behavioral demonstrations that four-dot OSM occurs at the whole-object level (i.e., after an object-level representation has been established) comes from the work by Lleras and Moore (2003), Moore and Lleras (2005). In their 2003 study, the mask and target were presented in different spatial locations and they manipulated the temporal dynamics of the display sequence. By doing this masking could be compared in situations where the dots and target were seen as separate objects with situations where the dots and target were seen as one object that moves and changes shape over time. Importantly, masking only occurred when the spatiotemporal sequence maintained the appearance of a *single* object that changed its appearance as it moved. Following this, Moore and Lleras (2005) further manipulated the likelihood that participants would see the target and mask as belonging to the same unitary object. In their experiments whenever the target was seen as a separate object from the mask, either because the two did not move together (Experiments 1 and 2) or because the target and mask color mismatched (Experiment 3), masking was reduced or eliminated[3].

[3] Using 3D displays, Andrea Lichtman and I have also shown that even in situations where the masking dots are represented in a separate object file from the target, visibility is disrupted only when the blank surface created by the dots is perceived as occluding the target shape (Kahan & Lichtman, 2006). Though this situation (where the mask and target are perceived as separate objects) may not be typical in experiments examining four-dot OSM, this effect must arise at the whole-object level since Andrea and I were able to rule out early image-based explanations of masking in our study.

In addition, using an adaptation technique and fMRI, Carlson, Rauschenberger, and Verstraten (2007) have pinpointed the lateral occipital cortex (LOC) as an area in the brain where four-dot OSM selectively exerts its influence. This is important because using this same brain-imaging approach, Kourtzi and Kanwisher (2001) have found that the LOC selectively responds to an object's overall perceived shape even when low-level contours differ, and this area does not respond to overlap in low-level contours when the overall perceived shape differs. In other words, this area responds to perception of the whole object rather than its lower-level parts. Taken together, the picture that has emerged is that four-dot OSM occurs at the whole-object level.

2.1.1 Summary of Key Findings

Four-dot OSM is the robust finding that target accuracy is impaired when a sparse four-dot mask, which surrounds a target, remains visible following target offset.

- Four-dot OSM occurs because of reentrant pathways in the visual system. This effect requires delayed offset of the masking dots and is measured as the difference in accuracy in the immediate and delayed-offset conditions (Di Lollo et al., 2000).
- Four-dot OSM arises relatively late in visual processing after an object-level representation has been established (i.e., at the whole-object level); this type of masking is larger when the mask and target are perceived as belonging to the same object file that changes over time (Lleras & Moore, 2003; Moore & Lleras, 2005) and may selectively involve processing in the LOC (Carlson et al., 2007).

2.2 Two-Dot Masking and Edge-Based Interactions at Early Stages of Vision

In 2002, Kathy Mathis and I conducted a series of experiments examining dot-masking effects and in the process we stumbled across an unusual result which we now refer to as object trimming (Kahan & Mathis, 2002). In Experiment 3 of that series we presented participants with two-dot masks that flanked target diamonds which were missing a corner on the left or right (see Figure 2). Our results indicated that when dots were arranged vertically and flanked the target's missing corner (Panel A in Figure 2) masking was absent but masking was robust when this same mask flanked the target's intact edge (Panel B in Figure 2). This led us to speculate that the mask could perceptually trim part of the nearby target, leading to an unambiguous

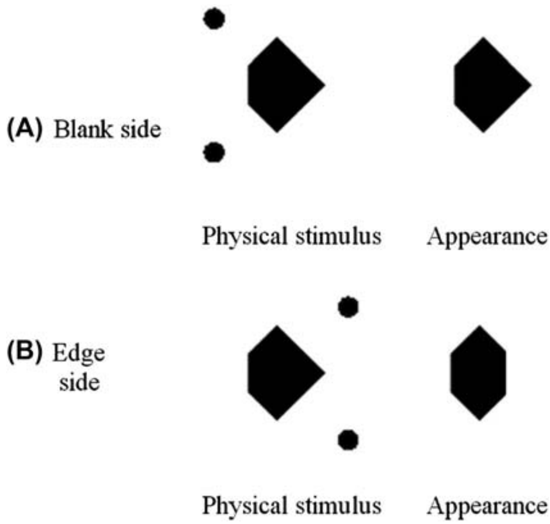

Figure 2 Depiction of the displays used in Kahan and Mathis (2002, Experiment 3) and the phenomenological appearance of these displays. In Panel A the dots flank the target's blank side and in Panel B the dots flank the target's intact edge. *This figure is based on a figure from Kahan and Mathis (2002), Psychonomic Society, and is adapted with permission.*

percept in situations where the dots flanked the blank side (see Panel A's appearance) and an ambiguous percept when the dots flank the target's edge side (see Panel B's appearance). However, since this result was tangential to the questions we were asking at that time we did not, in that report, pursue this possibility further.

2.2.1 Object Trimming

More recently, Jim Enns and I have conducted experiments systematically evaluating object trimming (Kahan & Enns, 2010, 2014). One of the key approaches we have taken when doing this, which allows us the capability of assessing both masking caused by reentrant processes and edge-based object trimming effects, is to have participants identify numbers and/or letters that appear in a digital-clock-style font. By doing this we can assess masking when the dots flank the target by measuring target identification accuracy when the dots have immediate relative to delayed offset (where the difference provides a measure of masking caused by reentrant processes) as well as examine object trimming directly. To measure object trimming we examine situations where people misidentify the target. When participants make a false alarm response to the target we can examine what, specifically, people report having seen. For example, if people are shown the target numerals 5, 6, 8, and 9, all in a digital-clock-style font and if these targets are all flanked by two dots that are arranged vertically on either the lower left or upper right, what will people report having seen when an error is made? For example, when people incorrectly identify the number 8 flanked by dots on the lower left (see Panel A in Figure 3), will they report having seen a 9 at a rate greater than expected by chance? Similarly, when people incorrectly identify the number 8 flanked by dots on the upper right (see Panel B in Figure 3), will they report having seen a 6 at a rate greater than expected by chance? We call these types of errors trimmed responses because these are the type of responses expected if the masking dots perceptually trim the nearby portion of the target (i.e., these are the numbers created by erasing the edge closest to the dots). Since there are three possible errors with four targets we reasoned that if people report trimmed responses at a rate greater than 33% of the errors then this would provide evidence that the dots do in fact alter perception of the target.

In all of our experiments we consistently find that masking from the adjacent dots (i.e., the difference in target accuracy in the immediate and delayed-offset conditions) is greater when the dots flank the intact edge (e.g., dots flanking the upper right of the number 9) relative to when the dots flank a blank side (e.g., dots flanking the lower left of the number 9) a result that replicates Kahan and Mathis (2002). More importantly, object trimming (i.e., reporting the trimmed response at rates greater than expected by chance) is robust and equivalent in both the immediate and delayed-offset conditions. That object trimming is unaffected by offset delay is important because it suggests that this effect, unlike four-dot OSM, is not

Figure 3 Depiction of the target numeral 8 (shown on left-hand side) flanked by dots on the lower left (Panel A) or the upper right (Panel B) and the appearance of these targets (shown on the right-hand side).

caused by reentrant processes. If object trimming had been caused by reentrant processing then more object trimming should have been found when the dots persisted on the screen after target offset. Since delayed offset is not needed for object trimming to occur, the data suggest that this arises during the initial feedforward pass. Specifically, object trimming occurs as a result of very early edge-based competition, and this conclusion is supported by several additional findings.

First, object trimming is critically dependent on the mask's illusory edge flanking the target's intact edge. We have found that when we insert a curved segment between the two dots to create a situation where the illusory edge no longer flanks the target's edge (see Panel A in Figure 4), trimming is

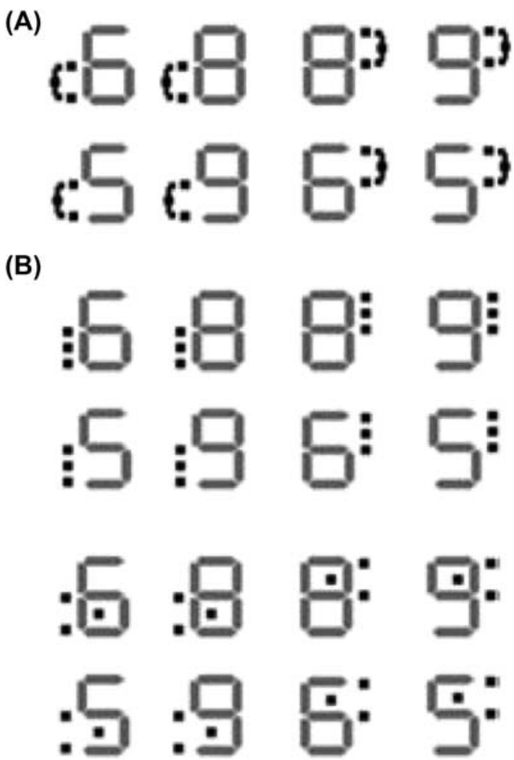

Figure 4 Depiction of targets and masking dots used in Kahan and Enns (2010). When a curved element is added between the two-dot mask (Panel A), object trimming is eliminated. When a third element is added that does not disrupt perception of the flanking edge (Panel B), object trimming is robust. *This figure is based on a similar figure from Kahan and Enns (2010), American Psychological Association, and is adapted with permission.*

completely eliminated while masking (i.e., the difference in accuracy between immediate and delayed offset) remains robust. However, when additional visual information that does not disrupt perception of the illusory edge is added (see Panel B in Figure 4) both trimming and masking remain intact, and trimming continues to be unaffected by offset delay. Taken together these data demonstrate that object trimming is caused by edge-based competition between the mask's illusory edge and the target's intact edge. When the target's edge and masking dots do not share an edge, no object trimming is found and when the target's edge and masking dots do share an edge object trimming is robust.

Second, object trimming disrupts perception before edges are assigned to an object-level representation. When we presented participants with the number 9 flanked by dots on the upper left (to create perception of a 3) or upper right (to create perception of a 5), we found that reaction times to identify a clearly presented target numeral as 3 or 5 were faster when the target was preceded by a prime that was trimmed to be congruent relative to when the prime was trimmed to be incongruent. Reaction times to identify the target numeral 3 are faster when preceded by a 9 with dots flanking the upper left relative to a 9 flanked by dots on the upper right and the opposite is found for the target numeral 5. Importantly, this repetition priming effect for targets preceded by primes that are trimmed to be congruent or incongruent was equivalent in magnitude to the priming obtained in situations where targets are preceded by primes that were actually identical (a 3 preceded by a 3) or mismatching (a 3 preceded by a 5). This finding indicates that object trimming occurs before edges have been assigned to whole objects. Consider the alternative hypothesis: if object trimming occurred only after edges were assigned to an object-level representation, then we would have expected to find stronger repetition priming when the prime and target actually matched one another. Instead we found equivalent priming when primes were trimmed to match the target.

Finally, object trimming arises before motion is perceived. This conclusion is based on the finding that the perceived direction of motion of a bistable motion quartet is influenced by object trimming. In bistable motion a person is presented with two objects that are located in opposite corners of an imaginary rectangle. This display is subsequently replaced with a second display where the objects appear in the previously unoccupied diagonal corners. When these displays are flickered back and forth, one after the other, motion is seen and the direction of motion (horizontal or vertical) is ambiguous and will vary from trial to trial (Ramachandran & Anstis, 1985). Using this approach we have found that when diamonds that are missing their left or right corner are presented in the opposite diagonals we can alter the perceived direction of motion (making horizontal or vertical motion more likely) by way of object trimming (see Figure 5). When people are shown the arrangement of events in Figure 5(A) and Frames 1 through 4 are cyclically repeated participants indicate that they perceive horizontal motion more often than is reported in situations where the arrangement of events matches Figure 5(B) (here vertical motion is more likely). This occurs because the dots trim the nearby corner of the intact diamond in Frame 3. As a result of this trimming effect, similarity of form influences the direction of perceived motion. This finding replicates others who have found that motion perception is influenced by similarity of form (Oyama, Simizu, & Tozawa, 1999), but here we accomplish this using object trimming. As such object trimming must occur before motion is perceived.

For the curious, examples of object trimming are available online at http://abacus.bates.edu/~tkahan/ot_demo.htm.

2.2.2 Object Binding

In addition to finding that two vertically aligned dots that flank a target will compete with that target's intact edge (object trimming), Jim Enns and I have found that this same two-dot mask when shown flanking a target's blank side can have a synergistic effect where the illusory edge created by the dots is falsely bound with the target, an effect we refer to as object binding (Kahan & Enns, 2014). This object binding effect has been studied in experiments using the same basic approach as those studies that examine object trimming. Namely, participants were asked to identify a briefly presented target that appeared in a digital-clock-style font. When this was done and the dot-mask flanked the lower right side of the letter P, for example, people reported having seen an A at a rate greater than chance when they misperceived the target (object binding; in the same way, if people are shown a 9 with dots flanking the lower left they will misperceive this as an 8). I note here that in this study we also introduced a new control technique for assessing both object trimming and object binding. In addition to comparing the rates of trimming and binding false alarm responses to chance levels, we also included a condition where target accuracy was reduced by decreasing the target's contrast. By doing this we could equate performance in the dot-mask condition with performance in the reduced contrast condition. One nice thing about this control condition is that by comparing

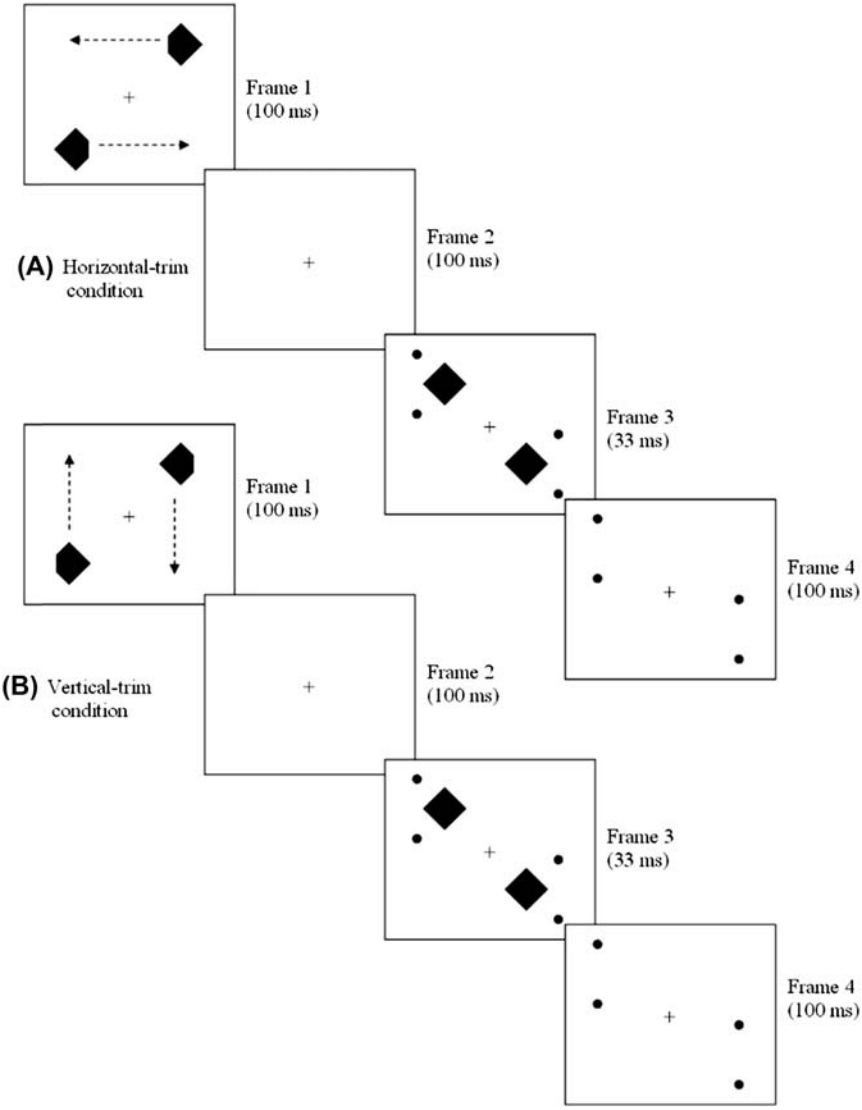

Figure 5 Depiction of the sequence of events and display types used in Kahan and Enns' (2010) Experiment 5. The perception of horizontal motion is more likely in Panel A and the perception of vertical motion is more likely in Panel B. Note that the dotted lines shown in Frame 1 did not appear in the actual displays that were presented to participants. See text for details. *This figure is taken from Kahan and Enns (2010), American Psychological Association, and is reprinted with permission.*

false alarm responses when the dots flank the target with the same false alarm responses in the control task, we can be certain that binding (or trimming) responses following an error are caused by the dots rather than reflecting any type of guessing bias when the target is difficult to see. Critically, when we compare false alarm responses with either chance levels or with this new control condition we find that edge-based interactions result in (1) competition when dots flank an intact edge (object trimming) and (2) cooperation when dots flank a target's blank side (object binding).

Of further importance, Kahan and Enns (2014) report that object binding is larger when the dots appear and disappear together with the target relative to when the dots have a delayed offset. The finding that object binding is larger in the immediate-offset condition than the delayed-offset condition indicates that (1) object binding, unlike object trimming, may be greater when the dots and target are seen as a holistic shape and this is more likely when the two have synchronous onsets and offsets (Alais, Blake, & Lee, 1998; Lee & Blake, 1999) and (2) object binding, like object trimming, arises during the initial feedforward pass and is not reliant on reentrant signals. If object binding had been caused by reentrant processing then we should have found greater object binding effects when the dots remained visible beyond target offset. Since object binding, like object trimming, is not larger with delayed offset (in fact object binding is reduced with delayed offset), this supports the conclusion that object binding arises during the initial feedforward pass.

2.2.3 Summary of Key Findings

Perception of a target's shape is altered when a sparse two-dot mask flanks a target's intact edge (causing object trimming) or blank side (causing object binding). Notably, object trimming and object binding differ in important ways from four-dot OSM because unlike four-dot OSM:

- Object trimming occurs during the initial feedforward sweep and does not require delayed offset of the masking dots; object trimming is measured as the conditional probability of making a "trimming" false alarm response when the target is falsely identified. Object trimming is robust and equivalent under immediate and delayed-offset conditions (Kahan & Enns, 2010, 2014).
- Object binding is larger in the immediate-relative to delayed-offset condition (Kahan & Enns, 2014), further supporting the position that edge-based interactions occur during the initial feedforward pass.

- Edge-based interactions arise early in visual processing before an object-level representation has been established; this is supported by the finding that performance is altered in tasks ranging from the identification of rapidly presented targets (Kahan & Enns, 2010, 2014), repetition priming (Kahan & Enns, 2010, Experiment 4), and perception of motion in ambiguous displays (Kahan & Enns, 2010, Experiment 5).

3. WHAT DOT-BASED MASKING CAN TELL US ABOUT COGNITION

As noted at the outset of this chapter, masking effects are quite often treated in perception textbooks as nothing more than an oddity that reflect constraints in perception which arise in artificial lab-based studies. This is unfortunate because it is now becoming clear that by leveraging what is known about dot-based masking effects we can learn a great deal about visual object recognition more broadly. In what follows I review recent experimental evidence in which dot-based masking is used as a tool for exploring the role of reentrant processing in awareness, the role of reentrant processing in feature binding, whether meaning is extracted during the initial feedforward pass in the visual system, whether meaning exerts a top-down influence on early stages of vision, and whether cultural expectations exert an influence on vision before objects are even recognized.

3.1 Tool for Examining Role of Reentry in Object Recognition

Objects are not static or unchanging, and reentrant processes assist in our maintenance and tracking of multiple objects across time and space (for a review of ways in which reentrant processes may contribute to the perception of other visual illusions, see Enns, Lleras, & Moore, 2010). But do these pathways also contribute to our conscious experience of seeing? And, if these pathways do contribute to consciousness, do these pathways help us to see object features as being unitary and bound with an object? By examining four-dot OSM, which is caused by activity along these pathways, there is growing evidence that reentry is essential to both our conscious experience of seeing objects as well as our ability to bind object features into a cohesive whole.

3.1.1 Is Reentry Needed for Awareness?

By using a four-dot OSM task and comparing performance with backward masking, Mika Koivisto (2012) has reached the conclusion that reentrant pathways are critical in helping people establish a clear and vivid conscious experience of visual objects. The claim that reentrant pathways may be necessary for awareness is not new (Bullier, 2001b; Hochstein & Ahissar, 2002; Lamme & Roelfsema, 2000); however, to date this has been a difficult topic to investigate experimentally. In the study by Koivisto, participants were either presented with a target item that was subsequently masked or they were presented with a blank display that was masked. The task was simply to determine if a target had been shown (present/absent decision). Not surprisingly, a signal detection analysis revealed that both four-dot OSM and backward masking[4] impaired signal sensitivity. However, in addition to examining detection accuracy with a signal detection analysis, Koivisto asked participants to rate their confidence in their response on a three-point scale (low, medium, or high). By comparing confidence ratings in situations where participants correctly responded that no target had been shown (i.e., correct rejection trials) with confidence ratings when participants incorrectly responded that no target had been shown (i.e., miss trials) and then contrasting these results for the two types of masking (backward masking and four-dot OSM) an interesting pattern emerged. When backward masking obscured vision, Koivisto found that confidence ratings in correct rejection and miss trials did not differ. This fits with the generally accepted idea that backward masking by a temporally near and overlapping pattern disrupts processing during the initial feedforward sweep, since this type of masking (which is also sometimes called integration masking) disrupts processing in the same set of neurons that respond to the target (Breitmeyer & Öğmen, 2006; Enns & Di Lollo, 2000). However, when four-dot OSM obscured vision the results were quite different. Here confidence ratings on correct rejection trials were higher than on miss trials even though participants reported not having seen a target on both of these trial types. Koivisto argues, and I agree, that the difference in confidence with four-dot OSM for miss and correct rejections may reflect continued activation in brain areas associated with low-level visual features on miss trials since four-dot OSM, unlike backward pattern masking, did not disrupt the feedforward pass. This is

[4] Backward masking involves presenting a distracting stimulus (visual noise) in the exact same spatial location as the target upon target offset.

important because it adds support for the proposal that the feedforward pass results in unconscious gist-level awareness of the visual scene and it is reentrant pathways that, if allowed to reexamine activity in low-level brain areas, support a more fine-grained and consciously rich experience of the visual world.

3.1.2 Is Reentry Needed for Feature Binding?

The visual world is filled with an abundant array of objects, each of which is composed of multiple visual features (color, form, movement, etc.) yet our visual processing system is not only able to extract and analyze these individual object features (Treisman & Gormican, 1988) but we manage to then recombine these features, which are represented in multiple brain regions, into a coherent and integrated percept. How this is done has been the source of controversy (Di Lollo, 2012; Treisman, 1996; Wolfe, 2012), and one possibility is that reentrant pathways aid in the binding of features (Tononi, Sporns, & Edelman, 1992).

To examine this issue, Bouvier and Treisman (2010) presented participants with colored bars that were either (1) horizontally or vertically positioned and (2) presented in isolation or with an accompanying and crossing white bar that had the opposite orientation. Two tasks were used. When participants' task was to indicate the color of the nonwhite bar no feature binding was necessary since only one nonwhite color appeared in the area of space defined by the four-dot mask on any trial and this was true both in situations where 2 bars overlapped and in situations where 1 bar appeared by itself. However, when participants' task involved identifying the orientation of the nonwhite bar, feature binding was necessary when 2 bars were shown since here there were two orientations (vertical and horizontal), and in order to determine the orientation of the nonwhite bar the color information needed to be bound with the form information. Using this task, Bouvier and Treisman only found four-dot OSM when binding of features was necessary (i.e., in conditions where 2 bars overlapped and the task involved identifying the orientation of the nonwhite bar). No four-dot OSM was found when binding was unnecessary (i.e., in conditions where 1 bar appeared by itself or in conditions where the task was to identify the color of the nonwhite bar). In addition, Bouvier and Treisman found that when a backward pattern mask was used to disrupt feedforward processing, masking was robust irrespective of whether binding was necessary or not. From this, Bouvier and Treisman made the argument that when feedforward processing is disrupted by a pattern mask performance in all conditions is low because visual feature extraction is impaired. More importantly though, when reentrant processing of target features is disrupted by four-dot OSM, performance is selectively impaired only in those situations where binding is necessary. These data support the position that visual feature binding requires reentry.

Following this, Goodhew, Edwards, Boal, and Bell (2015) questioned whether four-dot OSM only occurs when binding is necessary. In their experiments they found that four-dot OSM can occur even when binding is unnecessary so long as the target and mask look perceptually similar (e.g., both are similarly oriented gabor patterns) and are seen as the same object changing shape over time (see Lleras & Moore, 2003; Moore & Lleras, 2005 for similar conclusions). However, that four-dot OSM is influenced by target-mask visual similarity and can occur when binding is not necessary, does not, in itself, indicate that the reentrant processes that are used to update objects are not also used to bind together visual features.

In fact, in a follow-up study to Bouvier and Treisman (2010) that replicates this basic design, Koivisto and Silvanto (2011) report data from an additional dependent measure that adds convincing support for the claim that reentrant processes facilitate visual feature binding. In their studies, Koivisto and Silvanto replicate the basic task used by Bouvier and Treisman but in addition to examining target accuracy participants rated, on a trial-by-trial basis, whether they were (1) aware of both the color and the orientation of the stimulus, (2) aware of only the color but not the orientation, (3) aware of only the orientation but not the color, or (4) unaware of both the color and orientation. By doing this they were able to not only replicate the findings of Bouvier and Treisman (2010), which they did, but they were also able to examine the proportion of trials where illusory conjunctions were perceived. To do this, when two overlapping bars were shown and people correctly identified the color but incorrectly identified the orientation, Koivisto and Silvanto examined whether participants reported being aware of the orientation of the colored bar. If participants indicated that they were unaware of the bar's orientation the trial was coded as a nonillusory conjunction. However, if participants indicated that they were aware of the bar's orientation the trial was coded as an illusory conjunction, since the color information (which was correctly seen) was bound with the wrong orientation as evidenced by their reporting having seen the orientation of the colored bar. Koivisto and Silvanto (2011) found that nonillusory conjunctions are more likely when a backward masking pattern (which disrupts feedforward processing) is used than with four-dot OSM (which disrupts reentry) but illusory conjunctions show the reverse pattern. Illusory conjunctions were more likely with four-dot OSM (which disrupts reentry) than with a backward masking pattern (which disrupts feedforward processing). This result adds nice support for the conclusion that visual features are extracted during the first-wave feedforward pass, and that reentrant processes are needed in order to correctly bind these features together into a cohesive whole. If reentrant processes are disrupted then (1) participants have difficulty in the task when feature binding is necessary and (2) participants are more likely to falsely bind features together.

3.1.3 Summary of Key Findings

Reentrant pathways work to update the contents of object files over time but these pathways may also serve additional purposes including:

- Reentry appears to aid in the conscious experience of having seen an object. This conclusion receives preliminary support from the fact that participants are more confident in their perception of not having seen a target when nothing was shown relative to when they missed the target, provided that perceptual sensitivity is reduced by four-dot OSM; no difference in confidence is seen when perceptual sensitivity is reduced via backward pattern masking (Koivisto, 2012).
- Reentry facilitates the binding of visual features. This conclusion is supported by the fact that: (1) four-dot OSM has a larger influence on performance when identification of a target requires the binding of visual features (Bouvier & Treisman, 2010; Koivisto & Silvanto, 2011) and (2) illusory conjunctions are more likely when target visibility is reduced with four-dot OSM relative to backward pattern masking (Koivisto & Silvanto, 2011).

3.2 Tool for Investigating Meaning

The purpose of sight is to determine what information is present in the visual environment and people bring a lifetime of experiences and expectations that may help guide this processing. In this section I review experiments which have sought to determine whether the initial feedforward pass results in a semantically meaningful object-level representation. Following this I review experiments that seek to determine whether meaning exerts a top-down influence on the earliest stages of vision. To date, studies that have used dot-based masking to examine these questions suggest that gist-level and meaningful information *is* extracted during the initial feedforward pass, and that reentrant pathways *do* carry information about expectations and life experiences and this in turn influences the earliest stages of vision, before objects have even been identified.

3.2.1 Is Meaning Extracted by the Time the First-Wave Pass Is Complete?

The first-wave pass is fast acting[5] and after this initial processing sweep is completed, gist-level meaningful information may become available (Hochstein & Ahissar, 2002; Potter, 1975). However, that gist-level information is available for very briefly presented images does not, by itself, necessarily indicate that meaning was fully extracted during the bottom-up pass, since reentrant processes may have played some part in the extraction of meaning. To investigate whether meaning is extracted during the feedforward pass, a fine-grained analysis is needed and experiments assessing edge-based trimming and four-dot OSM may provide this.

3.2.1.1 Evidence from Four-Dot OSM

Several studies have examined whether targets that are rendered unconscious by four-dot OSM are nevertheless processed for semantic-level content. If feedforward processing of the target does not continue when the dots remain visible beyond target offset and the processing that occurs here is reentrant in nature only, then discovering how deeply the target was processed under these conditions would indicate the level of processing that was achieved by the end of the first-wave pass. However, it is certainly possible that bottom-up processing of the target continues after the target has been removed from the screen even when the dots remain visible in which case these studies may not indicate the exact level of processing achieved by the time the first-wave pass is complete. Either way, whether these studies tell us about the type of information extracted by the end of the first-wave pass, or during the time the dots remain visible, I turn to the question of whether a target that is rendered invisible by four-dot OSM is processed for meaning. This is important because it provides clues about the level of processing given to a target in four-dot OSM. For example, if the evidence suggests that targets are only processed for feature-based information, then the data would support there being a relatively cursory and low-level analysis of the masked item. However, if semantic-level content is available for targets that are masked with four-dot

[5] The exact speed of the first-wave pass likely depends on stimulus-level factors like contrast level, but this processing has been estimated as taking a mere 100 ms (see Lamme & Roelfsema, 2000).

OSM, this would suggest a much later locus of masking and a much richer analysis of the target.

In one of the first studies to tackle this issue, Woodman and Luck (2003) presented participants with target shapes that were masked using either four-dot OSM or backward pattern masking and recorded event-related potentials (ERPs). Their analysis focused on the N2pc component which is associated with the deployment of attention to an area of space that may contain task-relevant visual features (Kiss, Van Velzen, & Eimer, 2008). Results indicate that both four-dot OSM and backward pattern masking impair target visibility but only backward pattern masking, which impairs the feedforward sweep, disrupts the N2pc. The N2pc was unaffected by four-dot OSM. As such, it is clear that target-relevant features are extracted in four-dot OSM and that reentrant processing only disrupted awareness of the target's identity. But is meaning-based information, along with feature-based information, also extracted during the initial feedforward pass?

To answer this, Reiss and Hoffman (2006) conducted a similar study but rather than examining the N2pc component they focused on the N400 since this is associated with semantic-level processing (see Kutas & Hillyard, 1980, 1984). In their experiment, participants were shown two words, prime and target, which were either semantically related or unrelated to one another. The prime was always clearly shown and the target, which was surrounded by dots, was only clearly visible when the dots did not have a delayed offset[6]. Their N400 results indicate that although semantic-level processing is achieved when the dots terminate with the target, this processing is absent when dot offset was delayed. However, despite this clear result in the ERP data the behavioral data indicate that target accuracy was superior when the target was semantically related with the prime relative to when the prime and target were unrelated and this was true even for targets that were rendered invisible (suggesting that semantic-level processing was extracted after the bottom-up pass). One possible reason for this, as noted by Reiss and Hoffman (2006), was that when participants could not identify the target they may have guessed seeing a word related to the prime. So, when shown the prime word "saddle" participants might guess that the target was "horse." This type of guessing would subsequently increase

[6] In their experiment, targets were masked using eight small dots arranged in a rectangle, rather than the typical four-dot arrangement, but I note that the underlying mechanism responsible for impaired target processing in delayed-relative to immediate-offset conditions here is likely the same as that seen with four-dot masking (i.e., this impairment is caused by reentrant processes).

performance on masked targets that were related to the prime and would impair performance on unrelated masked targets. For this reason, the behavioral data are ambiguous while the ERP data support the conclusion that semantic-level processing is not achieved during the feedforward pass.

Following this, Reiss and Hoffman (2007) investigated the N170 component of the ERP waveform since this is associated with the recognition of faces and is larger when people are shown images of faces relative to other object categories like houses (Desjardins & Segalowitz, 2013). In Reiss and Hoffman's (2007) study, people viewed pictures of faces and houses that were surrounded by a dot mask, and offset delay was manipulated. Though the ERP data show that the difference in the N170 for faces and houses is robust in the immediate-offset condition and absent in the delayed-offset condition, the behavioral data, once again, are a bit problematic since participants could identify the target on 70% of the trials in the delayed-offset condition. As such, even though behavioral performance was impaired by offset delay (being worse in delayed—70%—relative to immediate—81%—offset conditions), it is surprising that the ERP data showed no difference between faces and houses in the delayed-offset condition when target visibility remained significantly above chance here. As such, it is possible that this electrophysiological measure did not fully tap into the categorical processing (face vs house) that did occur during the first-wave pass since the behavioral data clearly indicate that targets were correctly categorized on many trials, even when the dots had a delayed offset.

I note that there is an assumption in all of these electrophysiological studies that the waveform under scrutiny (whether it be the N2pc, the N400, or the N170) necessarily and fully taps into conscious as well as unconscious processes, and this assumption may not be accurate. In each of these studies the elimination of a specific waveform is taken as direct evidence that the type of processing associated with that component of the ERP waveform did *not* occur. It is clearly the case that if the ERP waveform associated with a certain level of processing is unperturbed by four-dot OSM then that processing must have continued in the delayed-offset condition despite the target being rendered unconscious. However, if the ERP waveform being examined is sensitive to a person's conscious experience of the target, and only occurs when people are conscious of seeing the target, then examinations of that ERP waveform are ill-suited for answering questions related to the type of processing that persists following four-dot OSM. This is not a trivial issue because several experiments suggest that both the N400 (Matsumoto, Iidaka, Nomura, & Ohira, 2005; Ruz, Madrid, Lupiáñez, & Tudela, 2003) and the N170 (Suzuki & Noguchi, 2013) are in fact affected by a person's conscious experience, and that these waveforms can be eliminated when processing is unconscious, even when behavioral evidence suggests processing did take place. As such, examinations of these waveforms may tell us very little, if anything, about the level of processing for targets that are rendered invisible by four-dot OSM (but see Kiefer, 2002 for evidence that the N400 does tap into unconscious processing). Future research should be careful of this by making every effort to measure waveforms that are equally sensitive to conscious and unconscious processes. One tentative possibility would be to examine the N200, rather than the N400, to the extent that this waveform more accurately indexes unconscious semantic-level processing (see Ruz et al., 2003).

It is for this reason that it may not be surprising that more convincing evidence now comes from behavioral experiments where the target and masking elements are either congruent or incongruent with one another, and reaction times to identify the masking elements are measured. For example, Chen and Treisman (2009) presented participants with four identical letters (e.g., E) arranged in a square that surrounded a target (e.g., the letter O, the letter F, or no target), and the surrounding masking letters either persisted beyond target offset or terminated with the target. Participants were required to make two responses. First, participants rapidly categorized the mask and following this participants indicated if a target was shown. Importantly, Chen and Treisman also manipulated the relationship between the target and masking items such that the two were either congruent or incongruent with one another at either a perceptual or categorical level. In their experiments, examining perceptual processing of the target the task was to categorize the mask as having straight or curved lines. When this was done the results indicate that irrespective of whether the target is seen or missed, reaction times to respond to the mask are faster when the target and mask are congruent than when they are incongruent. By way of example, participants are faster to indicate that a mask made up of four repeated E's has straight lines when this mask surrounded a letter that also had straight lines "F" relative to when this mask surrounded a letter with curved lines "O," and this was true even when the mask had a delayed offset and the target was missed. However, in their experiments examining categorical processing of the target the task was to categorize the mask as being a consonant or vowel. Here, reaction times to respond to the category of the mask were only affected by mask-target congruency when the target was seen. Again, by way of example, participants were faster to indicate that a mask made up of four repeated E's was a vowel when this mask surrounded a vowel "O" relative to a consonant "F," but here this pattern only emerged when the target was seen. No difference was found when the target was missed. This pattern of results supports the conclusion that the first-wave pass results in an object-level representation that is processed for basic visual features but that semantic-level information was not extracted from the stimulus.

However, Goodhew, Visser, Lipp, and Dux (2011) point out that the study reported by Chen and Treisman (2009) had one major limitation. Categorizing a letter as a vowel or consonant is not a highly practiced task for most individuals and as such it is likely that this type of semantic-level information is not automatically extracted from letters. Consequently, the failure to find an influence of mask-target congruency when targets are missed in their consonant—vowel categorization task may reflect poor task sensitivity rather than a lack of any semantic-level processing, per se. To better address this, Goodhew et al. presented participants with color words and noncolor words (both printed in a gray-colored font) that were masked using four-dot OSM. Importantly, the dots were also presented in a color and the color of the masking dots, and the target word sometimes matched (e.g., the word blue surrounded by blue dots) and sometimes mismatched (e.g., the word pink surrounded by blue dots). In their Experiment 1, they found that reaction times to respond to the color of the mask were faster when the word and dots were congruent relative to incongruent, and this was true no matter if the target word was identified correctly or if it was missed (here identification was determined using a nonspeeded lexical decision task). In their Experiment 2, they used a detection task, rather than a lexical decision task, to determine whether the masked item was seen. Here they again found that reaction times to respond to the color of the mask were faster when the word was congruent relative to incongruent when the item was seen, and the opposite was found when the item was missed. Negative priming for masked words is not new (e.g., Kahan, 2000), and here the finding of a significant negative compatibility effect (NCE) when the word could not be detected may reflect response inhibition (Eimer & Schlaghecken, 1998). Note that other mechanisms may also contribute to the NCE (see Kahan & Chokshi, 2013 for a brief overview of mechanisms that may contribute to the NCE). However, no matter what the cause of the NCE, these data bolster the claim that semantic information is extracted even when reentrant processing obscures the target since words that were not correctly identified (Experiment 1) or detected (Experiment 2) still had an influence on speeded reaction times to the mask.

One final point and a word of caution: it seems likely that the degree to which a stimulus activates semantic-level information when masked with four-dot OSM will depend on the type of stimulus shown (picture of an object, word, picture of a face, etc.). For example, in picture—word interference tasks when people are asked to categorize a word that is presented inside a congruent or incongruent picture, the congruency effect (faster reaction times on congruent relative to incongruent trials) is larger than in situations where people categorize the picture and ignore a congruent or incongruent word. Because the to-be-ignored picture interferes with the semantic categorization of a word more than a to-be-ignored word interferes with the semantic categorization of a picture, it has been suggested that pictures of objects activate semantic content more readily than words do (Smith & Magee, 1980; cf, Mathis, 2002). In addition, negative emotion-laden words may be processed more efficiently than neutral or positive emotion-laden words (Kahan & Hely, 2008) even when these words are not consciously seen (Prioli & Kahan, 2015), and emotion-laden faces may be processed even more efficiently than emotion-laden words (Beall & Herbert, 2008). It remains to be determined whether differences in the type of stimulus affects the level of activation achieved in four-dot OSM, but it seems likely that stimulus type along with low-level stimulus factors (size, contrast level, length of presentation, etc.) will play a role here; these possibilities await future testing.

Before it can be concluded, with certainty, that Goodhew et al. (2011) have resolved this issue and that the first-wave pass *does* results in semantic-level processing I point out that interference in the Stroop task may arise, to some extent, from nonsemantic response-level competition (Neely & Kahan, 2001). To illustrate what is meant by this, imagine that a participant is asked to press keys in response to meaningless symbols. Here the person might press key A for one symbol and key B for another symbol. Now imagine that the symbols are repeated to form masks, in a manner similar to what was done by Goodhew et al. (2011) and these meaningless masks surrounded these same meaningless symbols as targets. In a task of this sort it would not be surprising to see a flanker effect where reaction times responding to a masking symbol are faster when the target symbol is congruent relative to when the target symbol is incongruent. However, in this situation since the symbols are meaningless, this interference effect would most likely reflect response-level competition rather than interference at the semantic level. As such, better evidence for semantic-level processing would be to see differences in reaction times responding to colored masking dots in a replication of Goodhew et al. (2011) if the target words were a color-related term like "sky" (since this word is strongly associated with blue) relative to a noncolor-related word like "put" (matched on lexical characteristics) (see Klein, 1964). If a congruency effect emerged using this type of design (which Neely and Kahan call the sky/put design), then the data would unequivocally show semantic-level activation since the interference would require semantic-level processing of the masked word rather than simple response-level associations. This possibility awaits future testing.

3.2.1.2 Evidence from Object Trimming

Priming evidence shows that when a numeral is trimmed during the first-wave pass, this trimming culminates in a perceptual representation that is indistinguishable from a nontrimmed item. For example, in Kahan and Enns' (2010) Experiment 4 reaction times to the target numerals 3 and 5 were faster when preceded by a 9 that was trimmed to be congruent (e.g., a 9 flanked by dots on the upper left and upper right, respectively) with the target relative to when preceded by a 9 that was trimmed to be incongruent. In addition, the priming that was observed here was indistinguishable from priming to the targets 3 and 5 that were actually preceded by congruent or incongruent primes. This finding supports the claim that when edges are trimmed during the first-wave pass the resulting percept is eventually linked with semantic-level information, and this linking of the percept and semantics is unaffected by whether trimming had taken place at early stages of processing.

3.2.1.3 Summary of Key Findings

At this stage it is not fully clear whether semantic-level information is fully extracted during the initial bottom-up pass. Here I review key findings.

- Four-dot OSM: Several studies have examined whether the meaning of a target item is extracted after the bottom-up pass has run to completion, and the target is then rendered invisible by four-dot OSM. Here the results are somewhat mixed.
  - First, feature-level information is clearly extracted from unseen targets in four-dot OSM and this is supported by both electrophysiological studies (Woodman & Luck, 2003) as well as behavioral studies (Chen & Treisman, 2009).
  - Second, while some studies question whether semantic information is extracted from unseen targets in four-dot OSM, these studies have limitations.
    - Notably, electrophysiological studies were accompanied by behavioral data that undermine these conclusions (Reiss & Hoffman, 2006, 2007). In addition these studies make the critical assumption

that the elimination of an ERP waveform means that the process associated with this waveform never occurred and this assumption is not fully supported (Matsumoto et al., 2005; Ruz et al., 2003; Suzuki & Noguchi, 2013).

  – The behavioral data for there not being semantic-level activation in four-dot OSM may suffer from reduced task sensitivity (Chen & Treisman, 2009). As Goodhew et al. (2011) correctly point out the labeling of letters as being a consonant or vowel is not well practiced and automatized and as such the study by Chen and Treisman (2009) may have been incapable of detecting semantic-level processing.

- Finally, the most promising evidence comes from the study by Goodhew et al. (2011) which suggests that meaning *is* extracted from targets that are rendered invisible by four-dot OSM. What is clear from these data is that the first-wave pass results in the formation of an object-level representation that is fully linked with a behavioral response. However, the data do not yet unequivocally support there being full-blown semantic-level activation. This possibility awaits further testing and one approach, suggested here, is to combine Goodhew et al.'s (2011) paradigm with the sky/put design described by Neely and Kahan (2001).

- Object trimming: One study has examined the nature of the object-level representation that arises following object trimming (Kahan & Enns, 2010; Experiment 4). Here, object trimming (for the number 9) resulted in the formation of an object-level representation that was linked with a numeric label (3 or 5). Importantly, this representation was indistinguishable from the representation of a number that was not trimmed during the first-wave pass (i.e., both trimmed and untrimmed primes resulted in equivalent priming effects). This result supports the position that feature-level interactions take place during the first-wave pass, and that these early edge-based interactions influence the object-level representation that gets linked with long-term memory and becomes available to consciousness.

3.2.2 Does Meaning Exert a Top-Down Influence on Object Recognition?

It is well established that meaningful forms may be easier to see and harder to mask than meaningless shapes (Elze, Song, Stollhoff, & Jost, 2011; Reicher, 1969; Shelley-Tremblay & Mack, 1999; Weisstein & Harris, 1974; Wheeler, 1970), but does this difference in visibility for denotive and nondenotive shapes come about only after an object-level representation has been formed and linked with the contents of semantic memory or can the same reentrant pathways that are used to enhance processing of the stimulus carry information that guides the formation of objects, even before that object has been recognized? Up until now, there has been no way of assessing whether meaning has a top-down influence on the earliest stages of vision. What makes this theoretical puzzle solvable, for the first time, is the discovery of "object trimming" and "object binding," which arise during the first-wave pass of information processing. By comparing the effect that meaning has on these edge-based masking effects with the effects of meaning on masking at the whole-object level, it is now possible to disentangle the influence, if any, of higher-level cognitive processes on early stages of vision.

3.2.2.1 Evidence from Meaningful Letters and Meaningless Shapes

In 2014, Jim Enns and I compared object-based (four-dot OSM) and edge-based (object trimming and object binding) masking effects for targets that were either familiar and meaningful or unfamiliar and meaningless (see Figure 6).

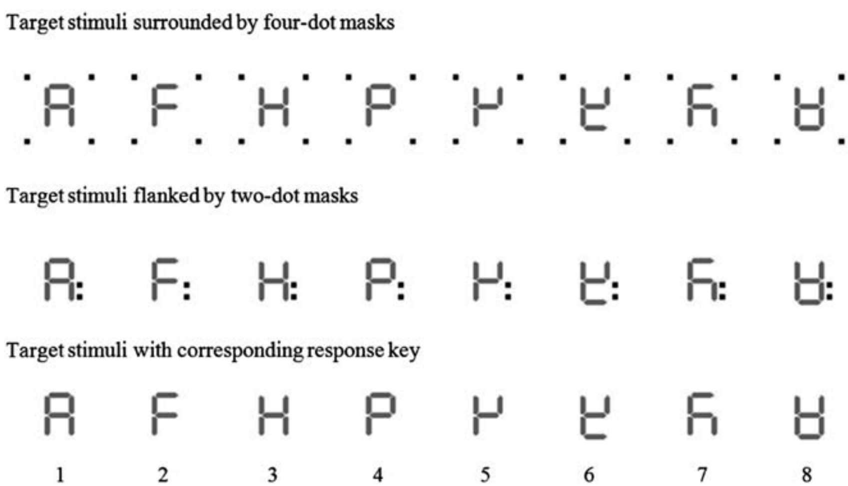

Figure 6 Examples of target stimuli used in Kahan and Enns (2014). Targets were either surrounded by four dots (top panel) or were flanked on the lower right by two dots (middle panel) and were either familiar forms (left-hand side) or unfamiliar shapes (right-hand side). The response key (bottom panel) was provided following target offset and participants used the numeric keypad to indicate the identity of the target. *This figure is taken from Kahan and Enns (2014), American Psychological Association, and is reprinted with permission.*

The critical question was whether shapes which are linked to a long-term memory representation (i.e., the familiar shapes shown on the left-hand side of Figure 6) are less susceptible to object-based and edge-based masking effects than shapes which are not linked to long-term memory representations (i.e., the unfamiliar shapes shown on the right-hand side of Figure 6). We hypothesized that if meaning only exerts its influence after the feedforward sweep has been completed, then familiar letters may be less susceptible to four-dot OSM than unfamiliar shapes, since four-dot OSM is a late-acting effect. Nevertheless, edge-based masking may be unaffected by meaning, since object trimming and object binding are early-acting effects. However, if meaning exerts a top-down influence before edges are assigned to objects, then we hypothesized an influence of meaning for both four-dot OSM and edge-based trimming and binding effects. Targets and masks were rapidly shown, and following this participants were provided with a response key (bottom portion of Figure 6) for identifying the target.

Our results clearly indicated that unfamiliar shapes were more susceptible to four-dot OSM, object trimming, and object binding. In addition, we ruled out the possibility that the trimming and binding effects were arising from guessing strategies by comparing the false alarm responses for targets that were flanked by two dots with the false alarm responses for targets that were made difficult to identify because their contrast was reduced (our control condition). This comparison clearly indicated that our results were not caused by guessing strategies but that perception of the target's form was altered by the flanking dots. As shown in Figure 7, people responded with trimmed false alarm responses (Panel A of Figure 7) and binding false alarm responses (Panel B of Figure 7) at rates much greater than seen in the control condition, and this was more likely when the target was unfamiliar (right side of both panels in Figure 7) relative to when the target was familiar (left side of both panels in Figure 7).

3.2.2.2 Evidence from Cross-Cultural Comparisons

Despite the fact that the broad characteristics of visual physiology are invariant across people from different cultures, perception researchers now generally *dis*agree with the long-standing dogma that people's perceptual experiences faithfully represent the visual world, and that these experiences are the same across viewers. In fact, visual illusions make it clear that perception is not always veridical and a relatively small, but growing, body of research studies indicate that people with different life experiences and from different cultures may see the world in different ways.

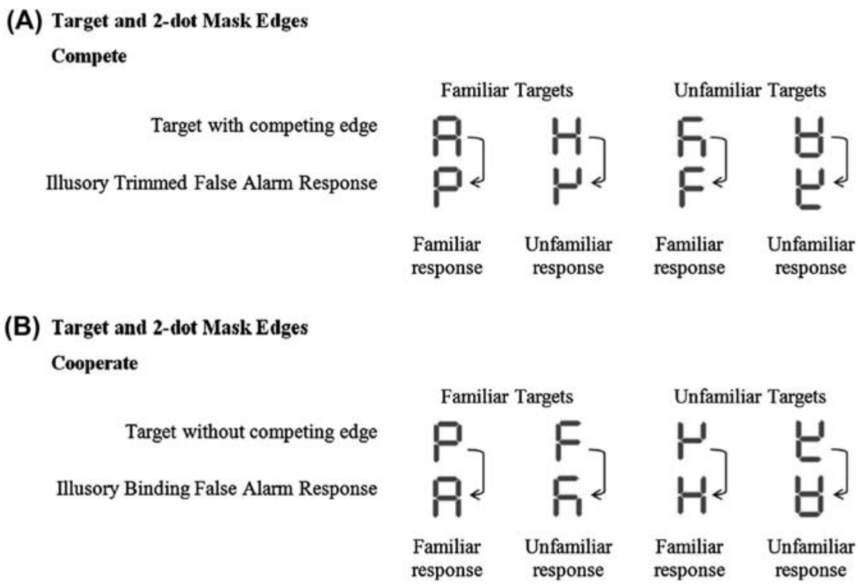

Figure 7 Depiction of targets and false alarm responses from Kahan and Enns (2014). When two dots flanked the target they always flanked the target's bottom right side. Panel A depicts situations where masking dots compete with the target edge (object trimming) and Panel B depicts situation where masking dots cooperate (object binding). The left-hand side shows familiar targets and the right-hand side shows unfamiliar targets. *This figure is taken from Kahan and Enns (2014), American Psychological Association, and is reprinted with permission.*

As one example of how culture affects vision, Masuda and colleagues have found that East Asian participants attend more to the context in which visual objects appear than Westerners (Masuda & Nisbett, 2001, 2006). For example, when people are asked to describe previously shown video clips, East Asian participants are more likely to describe contextual and relational information than are Westerners (Masuda & Nisbett, 2001). Similarly, when making speeded recognition judgments ("yes" vs "no") regarding whether a picture of an object was previously seen, East Asian participants are more accurate when the object was presented in the same context as it had appeared when first presented, whereas Western participants are not strongly affected by this type of contextual manipulation (Masuda & Nisbett, 2001).

Change blindness is also affected by culture. Change blindness is the robust finding that when people are shown two, slightly altered images (A and A′) with a delay between each, it takes a substantial period of time and continued flickering of the two images (A, A′, A, A′,…) before the change is successfully identified (Rensink, O'Regan, & Clark, 1997;

Simons, 2000). In a landmark study examining the effects of culture on change blindness, Masuda and Nisbett (2006) found that American participants detect changes in focal objects more quickly than changes in contextual (or background) information, but Japanese participants detect changes in focal objects and contextual information equally well, and Japanese participants detect changes in contextual information more quickly than American participants. In addition, Hohyung Choi, Cherelle Connor, Sarah Wason, and I have found that priming interdependent relative to independent self-knowledge can also influence the way people attend to visual scenes (see Choi, Connor, Wason, & Kahan, 2015). Prior work has shown that the degree to which someone views the self in more independent or interdependent terms varies across cultures, being more independent in Western cultures and interdependent in Eastern cultures (Markus & Kitayama, 1991). We found that Western participants were faster at identifying changes in focal objects relative to contextual objects in a change blindness task modeled after Masuda and Nisbett (2006), but that this difference in reaction times was reduced in half for Western participants who were primed with interdependency relative to Western participants primed with independence.

As several more examples, the Müller-Lyer illusion, which is the finding that people misperceive a line as being longer when it is surrounded by outward facing angles (e.g., >———<) than when it is surrounded by inward facing angles (e.g., <———>), differs dramatically across cultures. This illusion does not occur among the San of the Kalahari (Segall, Campbell, & Herskovits, 1966; see also Henrich, Heine, & Norenzayan, 2010). In addition, Turnbull (1961) reported that a 22-year-old BaMbuti man who had grown up in the rain forest of the Congo and as such had no experience with seeing objects from a great distance (since trees would normally obstruct their view), misperceived distant objects, like a herd of buffalo, as being the size of insects, rather than being far away. It is possible that a lifetime of seeing objects at a close distance affects one's ability to experience size constancy. For example, when children are asked to judge the size of coins with different same-sized circles, children from poorer areas rate the coins as being larger than do children from more affluent families (Bruner & Goodman, 1947), which illustrates that visual perception is influenced by a person's subcultural experience within the same society. Importantly, in all of these prior studies, participants have been able to view the visual information for a prolonged period before a response was required. As such it has not yet been established when in the visual processing stream culture exerts its influence. Does culture, like other types of meaning, exert a top-down influence on the edge-assembly process, even before objects are recognized?

To answer this question, Melody Altschuler, Paul Dux, and I have manipulated cultural expectations by presenting participants from the United States and Australia with images of either Abraham Lincoln or Queen Elizabeth II as they are displayed on the $5.00 bill from each country, respectively (Kahan, Altschuler, & Dux, 2015). The key question we were interested in answering was whether edge-based masking (object trimming) would be influenced by culturally mediated expectations. If object trimming, which occurs during the feedforward sweep, is affected by cultural expectations then this would be evidence for culturally learned information influencing the process of edge assignment to objects. For example, would American participants be more likely than Australian participants to trim a 9 into a 5 when the target is flanked by dots on the upper right and an image of Lincoln is shown on the screen (see Panel A in Figure 8) and would Australian participants be more likely to trim a 9 into a 5 than American participants if the image had been of Queen Elizabeth II (Panel B in Figure 8).

In this experiment, we examined performance in two phases. In the first phase target numerals were flanked by dots, and participants were required to identify the target shown (experimental). In the second phase no dots were shown, and target accuracy was matched to the average accuracy in Phase 1, for each subject (control). This was achieved by presenting targets in gray on a white background and the contrast level of targets in Phase 2 was adjusted on a trial-by-trial basis so that it was made lighter or darker depending on the participant's performance. Object trimming effects varied as a function of the face type (Lincoln or Queen Elizabeth), the participant's country of origin (Australia vs United States), and whether dots flanked the target (Phase 1; experimental) or whether the target was made difficult to identify by decreasing its contrast (Phase 2; control) (i.e., there was a three-way interaction). What this interaction indicates is that trimming of a 9 into a 5 was more likely for US participants when Lincoln appeared in the center of the screen (i.e., more trimming in Figure 8(A) relative to Figure 8(B)) but the opposite pattern emerged for participants from Australia (more trimming in Figure 8(B) than Figure 8(A)), and this two-way crossover interaction was limited to Phase 1 where dots flanked the target. In Phase 2 (control) when errors were made, participants made trimming false alarm responses at rates that were significantly lower than in Phase 1 (experimental) and here, unlike when dots flanked the target, trimming was not affected by the face type and the participant's country of origin. These

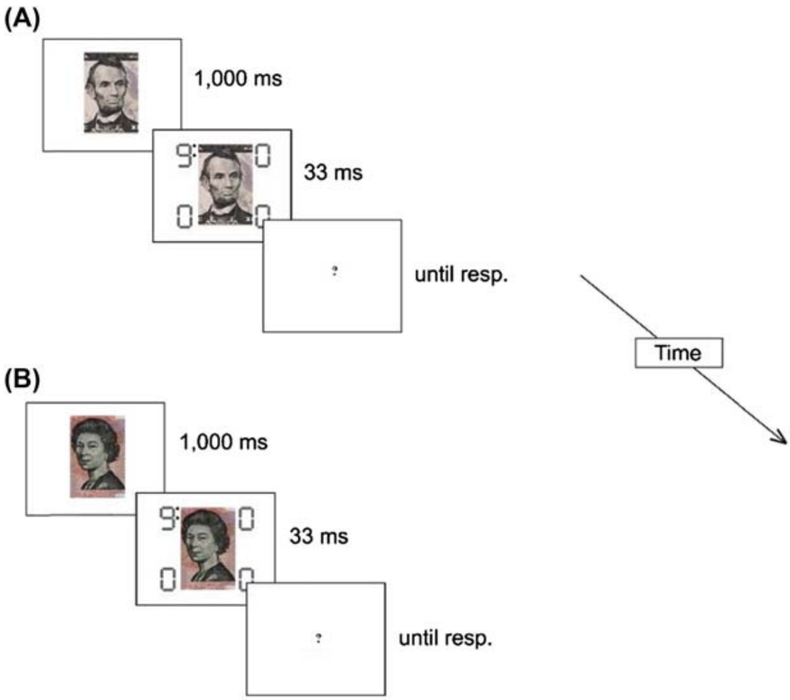

Figure 8 Sequence of events and conditions from Kahan et al. (2015). In Panel A the image of Abraham Lincoln as depicted on the US $5.00 bill appears centered on the display and in Panel B the image of Queen Elizabeth II as depicted on the Australian $5.00 bill appears centered on the display.

data clearly indicate that (1) there is a top-down influence of culturally mediated expectations during the first-wave pass of visual information processing and (2) this top-down influence occurs during the edge-assembly process, and before the object has even been recognized.

This result is important not only because it firmly establishes an effect of culture at early stages of vision but also because it clarifies an issue that was not fully addressed by Kahan and Enns (2014). In Kahan and Enns' (2014) study, critics could argue that rather than there being a top-down influence of meaning on early stages of perception, meaningful shapes may have more well entrenched and practiced bottom-up pathways, and it may be this benefit in bottom-up processing rather than a top-down influence that makes meaningful shapes less susceptible to edge-based masking. Here, the numbers 9 and 5 are equally meaningful to participants from the United States and Australia and as such these numbers should have equivalent bottom-up processing. In addition, since the likelihood of trimming a 9 into a 5 varied as a function of the face shown and the person's country of origin, and since this was not caused by guessing when the target was difficult to identify, the data must reflect an influence of top-down culturally mediated expectations on the edge-assembly process.

3.2.2.3 Summary of Key Findings

The same reentrant pathways that give rise to four-dot OSM and help with feature binding and conscious awareness may also carry information about meaning, expectations, and life experiences and this, in turn, guides the initial feedforward pass. This conclusion is supported by the finding that:

- Familiar shapes are less susceptible to four-dot OSM, and edge-based object trimming and object binding effects than are meaningless forms (Kahan & Enns, 2014).
- Culturally mediated expectations influence the likelihood that a person's visual system will trim off the nearby edge from a target numeral and this effect does not reflect guessing biases since there was no influence of the face (Lincoln or Queen Elizabeth) on perception in the control phase where dots did not flank the target (Kahan et al., 2015).

4. CONCLUSIONS

Though the iterative exchange of visual information can, in principle, make it difficult to pin down where in the visual system variables are exerting their influence, I believe that by studying masking at the whole-object and edge-based levels researchers are gaining better insights into this dynamic process. This is an important development. It is an exciting time to be a vision scientist and by tapping into what is known about object-based and edge-based masking effects researchers will undoubtedly gain a more solid understanding of how visual processing unfolds. Currently, the data from dot-masking experiments indicate that reentrant processes may aid in the conscious experience of seeing an object, that reentrant processes may be critical in the binding of object features, that the first-wave bottom-up pass results in the formation of an object-level representation that is linked with its response and may also activate semantic-level codes, and that reentrant processes carry information about meaning to low-level areas and as such meaning, and cultural expectations, help guide the edge-assembly process before objects have even been recognized.

ACKNOWLEDGMENTS

This work was generously completed with support provided by a James McKeen Cattell Fund Fellowship. I thank Jim Enns, Kathy Mathis, and Brian Ross for helpful comments. Correspondence should be addressed to Todd A. Kahan, PhD, Department of Psychology, Bates College, 4 Andrews Road, Lewiston, ME 04240. E-mail: tkahan@bates.edu.

REFERENCES

Alais, D., Blake, R., & Lee, S. H. (1998). Visual features that vary together over time group together over space. *Nature Neuroscience, 1*, 160–164. http://dx.doi.org/10.1038/1151.

Beall, P. M., & Herbert, A. M. (2008). The face wins: stronger automatic processing of affect in facial expressions than words in a modified Stroop task. *Cognition and Emotion, 22*(8), 1613–1642. http://dx.doi.org/10.1080/02699930801940370.

Bouvier, S., & Treisman, A. (2010). Visual feature binding requires reentry. *Psychological Science, 21*(2), 200–204. http://dx.doi.org/10.1177/0956797609357858.

Breitmeyer, B. G., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision* (2nd ed.). New York, NY, USA: Oxford University Press.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *The Journal of Abnormal and Social Psychology, 42*(1), 33–44. http://dx.doi.org/10.1037/h0058484.

Bullier, J. (2001a). Integrated model of visual processing. *Brain Research Reviews, 36*(2–3), 96–107. http://dx.doi.org/10.1016/S0165-0173(01)00085-6.

Bullier, J. (2001b). Feedback connections and conscious vision. *Trends in Cognitive Sciences, 5*(9), 369–370. http://dx.doi.org/10.1016/S1364-6613%2800%2901730-7.

Bullier, J., McCourt, M. E., & Henry, G. H. (1988). Physiological studies on the feedback connection to the striate cortex from cortical areas 18 and 19 of the cat. *Experimental Brain Research, 70*, 90–98.

Carlson, T. A., Rauschenberger, R., & Verstraten, F. J. (2007). No representation without awareness in the lateral occipital cortex. *Psychological Science, 18*(4), 298–302. http://dx.doi.org/10.1111/j.1467-9280.2007.01892.x.

Chen, Z., & Treisman, A. (2009). Implicit perception and level of processing in object-substitution masking. *Psychological Science, 20*(5), 560–567. http://dx.doi.org/10.1111/j.1467-9280.2009.02328.x.

Choi, H., Connor, C. B., Wason, S. E., & Kahan, T. A. (2015). The effects of interdependent and independent priming on Western participants' ability to perceive changes in visual scenes. *Journal of Cross-Cultural Psychology*. http://dx.doi.org/10.1177/0022022115605384. Published online before print September 13, 2015.

Desjardins, J. A., & Segalowitz, S. J. (2013). Deconstructing the early visual electrocortical responses to face and house stimuli. *Journal of Vision, 13*(5). http://dx.doi.org/10.1167/13.5.22.

Di Lollo, V. (2012). The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences, 16*(6), 317–321.

Di Lollo, V. (2014). Reentrant processing mediates object substitution masking: comment on Põder (2013). *Frontiers in Psychology, 5*.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*, 481–507.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2002). Object substitution without reentry? *Journal of Experimental Psychology: General, 131*(4), 594–596. http://dx.doi.org/10.1037/0096-3445.131.4.594.

Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1737–1747. http://dx.doi.org/10.1037/0096-1523.24.6.1737.

Elze, T., Song, C., Stollhoff, R., & Jost, J. (2011). Chinese characters reveal impacts of prior experience on very early stages of perception. *BMC Neuroscience, 12*. http://dx.doi.org/10.1186/1471-2202-12-14.

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research, 44*, 1321–1331.

Enns, J. T. (2008). Object substitution masking. *Scholarpedia, 3*(2), 3329. http://dx.doi.org/10.4249/scholarpedia.3329.

Enns, J. T., & Di Lollo, V. (1997). Object substitution: a new form of masking in unattended visual locations. *Psychological Science, 8*(2), 135–139. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00696.x.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*(9), 345–352. http://dx.doi.org/10.1016/S1364-6613(00)01520-5.

Enns, J. T., Lleras, A., & Di Lollo, V. (2006). A reentrant view of visual masking, object substitution, and response priming. In H. Öğmen, & B. G. Breitmeyer (Eds.), *The first half second: The microgenesis and temporal dynamics of unconscious and conscious visual processes* (pp. 127–147). Cambridge, MA, USA: MIT Press.

Enns, J. T., Lleras, A., & Moore, C. M. (2010). Object updating: a force for perceptual continuity and scene stability in human vision. In R. Nijhawan (Ed.), *Problems of space and time in perception and action* (pp. 503–520). Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511750540.028.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in primate visual cortex. *Cerebral Cortex, 1*, 1–47.

Francis, G., & Hermens, F. (2002). Comment on "Competition for consciousness among visual events: the psychophysics of reentrant visual processes" (Di Lollo, Enns, & Rensink, 2000). *Journal of Experimental Psychology: General, 131*, 590–593.

Goodhew, S. C., Edwards, M., Boal, H. L., & Bell, J. (2015). Two objects or one? Similarity rather than complexity determines objecthood when resolving dynamic input. *Journal of Experimental Psychology: Human Perception and Performance, 41*(1), 102–110. http://dx.doi.org/10.1037/xhp0000022.

Goodhew, S. C., Pratt, J., Dux, P. E., & Ferber, S. (2013). Substituting objects from consciousness: a review of object substitution masking. *Psychonomic Bulletin and Review, 20*(5), 859–877. http://dx.doi.org/10.3758/s13423-013-0400-9.

Goodhew, S. C., Visser, T. A. W., Lipp, O. V., & Dux, P. E. (2011). Competing for consciousness: prolonged mask exposure reduces object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 588–596. http://dx.doi.org/10.1037/a0018740.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. http://dx.doi.org/10.1017/S0140525X0999152X.

Hirose, N., Kihara, K., Mima, T., Ueki, Y., Fukuyama, H., & Osaka, N. (2007). Recovery from object substitution masking induced by transient suppression of visual motion processing: a repetitive transcranial magnetic stimulation study. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1495–1503. http://dx.doi.org/10.1037/0096-1523.33.6.1495.

Hirose, N., Kihara, K., Tsubomi, H., Mima, T., Ueki, Y., Fukuyama, H., et al. (2005). Involvement of V5/MT in object substitution masking: evidence from repetitive transcranial magnetic stimulation. *Neuroreport, 16*, 491–494.

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804.

Kahan, T. A. (2000). Negative priming from masked words: retrospective prime clarification or center-surround inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1392–1410. http://dx.doi.org/10.1037/0278-7393.26.6.1392.

Kahan, T.A., Altschuler, M.R., & Dux, P.E. (2015). Culturally-mediated expectations influence perception before object recognition. (Unpublished manuscript).

Kahan, T. A., & Chokshi, N. S. (2013). Perceptual interactions between primes, masks, and targets: further evidence for object updating. *Visual Cognition, 21*(1), 123–138. http://dx.doi.org/10.1080/13506285.2013.788112.

Kahan, T. A., & Enns, J. T. (2010). Object trimming: when masking dots alter rather than replace target representations. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 88–102. http://dx.doi.org/10.1037/a0016466.

Kahan, T. A., & Enns, J. T. (2014). Long-term memory representations influence perception before edges are assigned to objects. *Journal of Experimental Psychology: General, 143*(2), 566–574. http://dx.doi.org/10.1037/a0033723.

Kahan, T. A., & Hely, C. D. (2008). The role of valence and frequency in the emotional Stroop task. *Psychonomic Bulletin and Review, 15*(5), 956–960. http://dx.doi.org/10.3758/PBR.15.5.956.

Kahan, T. A., & Lichtman, A. S. (2006). Looking at object-substitution masking in depth and motion: toward a two-object theory of object substitution. *Perception and Psychophysics, 68*(3), 437–446. http://dx.doi.org/10.3758/BF03193688.

Kahan, T. A., & Mathis, K. M. (2002). Gestalt grouping and common onset masking. *Perception and Psychophysics, 64*, 1248–1259.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, & D. A. Davies (Eds.), *Varieties of attention*. New York: Academic Press.

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research, 13*(1), 27–39. http://dx.doi.org/10.1016/S0926-6410(01)00085-4.

Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology, 45*(2), 240–249. http://dx.doi.org/10.1111/j.1469-8986.2007.00611.x.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology,* 77(4), 576–588. http://dx.doi.org/10.2307/1420768.

Koivisto, M. (2012). Is reentry critical for visual awareness of object presence? *Vision Research, 63*, 43–49. http://dx.doi.org/10.1016/j.visres.2012.05.001.

Koivisto, M., & Silvanto, J. (2011). Relationship between visual binding, reentry and awareness. *Consciousness and Cognition: An International Journal, 20*(4), 1293–1303. http://dx.doi.org/10.1016/j.concog.2011.02.008.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*, 1506–1509.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205. http://dx.doi.org/10.1126/science.7350657.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163. http://dx.doi.org/10.1038/307161a0.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571–579.

Lee, S., & Blake, R. (1999). Visual form created solely from temporal structure. *Science, 284*(5417), 1165–1168. http://dx.doi.org/10.1126/science.284.5417.1165.

Lleras, A., & Moore, C. M. (2003). When the target becomes the mask: using apparent motion to isolate the object-level component of object substitution masking. *Journal of*

Experimental Psychology: Human Perception and Performance, 29, 106–120. http://dx.doi.org/10.1037/0096-1523.29.1.
Markus, H. R., & Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. http://dx.doi.org/10.1037/0033-295X.98.2.224.
Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology, 81*(5), 922–934. http://dx.doi.org/10.1037/0022-3514.81.5.922.
Masuda, T., & Nisbett, R. E. (2006). Culture and change blindness. *Cognitive Science, 30*(2), 381–399. http://dx.doi.org/10.1207/s15516709cog0000_63.
Mathis, K. M. (2002). Semantic interference from objects both in and out of a scene context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 171–182. http://dx.doi.org/10.1037/0278-7393.28.1.171.
Matsumoto, A., Iidaka, T., Nomura, M., & Ohira, H. (2005). Dissociation of conscious and unconscious repetition priming effect on event-related potentials. *Neuropsychologia, 43*(8), 1168–1176. http://dx.doi.org/10.1016/j.neuropsychologia.2004.11.020.
Mignard, M., & Malpeli, J. G. (1991). Paths of information flow through visual cortex. *Science, 251*, 1249–1251.
Moore, C. M., & Lleras, A. (2005). On the role of object representations in substitution masking. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1171–1180. http://dx.doi.org/10.1037/0096-1523.31.6.
Mumford, D. (1991). On the computational architecture of the neocortex I. The role of the thalamo-cortical loop. *Biological Cybernetics, 65*, 135–145.
Mumford, D. (1992). On the computational architecture of the neocortex II. The role of cortico-cortical loops. *Biological Cybernetics, 66*, 241–251.
Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger, J. S. Nairne, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 69–93). Washington, DC: American Psychological Association.
Oyama, T., Simizu, M., & Tozawa, J. (1999). Effects of similarity on apparent motion and perceptual grouping. *Perception, 28*, 739–748.
Põder, E. (2013). Attentional gating models of object substitution masking. *Journal of Experimental Psychology: General, 142*(4), 1130–1141. http://dx.doi.org/10.1037/a0030575.
Potter, M. C. (1975). Meaning in visual scenes. *Science, 187*, 965–966.
Prioli, S. C., & Kahan, T. A. (2015). Identifying words that emerge into consciousness: effects of word valence and unconscious previewing. *Consciousness and Cognition, 35*, 88–97. http://dx.doi.org/10.1016/j.concog.2015.04.005.
Ramachandran, V. S., & Anstis, S. M. (1985). Perceptual organization in multistable apparent motion. *Perception, 14*, 135–143.
Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 275–280. http://dx.doi.org/10.1037/h0027768.
Reiss, J. E., & Hoffman, J. E. (2006). Object substitution masking interferes with semantic processing: evidence from event-related potentials. *Psychological Science, 17*(12), 1015–1020. http://dx.doi.org/10.1111/j.1467-9280.2006.01820.x.
Reiss, J. E., & Hoffman, J. E. (2007). Disruption of early face recognition processes by object substitution masking. *Visual Cognition, 15*(7), 789–798. http://dx.doi.org/10.1080/13506280701307035.
Rensink, R. A., O'Regan, J., & 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. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00427.x.

Ruz, M., Madrid, E., Lupiáñez, J., & Tudela, P. (2003). High density ERP indices of conscious and unconscious semantic priming. *Cognitive Brain Research, 17*(3), 719–731. http://dx.doi.org/10.1016/S0926-6410(03)00197-6.
Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception*. Oxford, England: Bobbs-Merrill.
Shelley-Tremblay, J., & Mack, A. (1999). Metacontrast masking and attention. *Psychological Science, 10*, 508–515. http://dx.doi.org/10.1111/1467-9280.00197.
Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7*(1–3), 1–15. http://dx.doi.org/10.1080/135062800394658.
Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture–word processing. *Journal of Experimental Psychology: General, 109*(4), 373–392. http://dx.doi.org/10.1037/0096-3445.109.4.373.
Suzuki, M., & Noguchi, Y. (2013). Reversal of the face-inversion effect in N170 under unconscious visual processing. *Neuropsychologia, 51*(3), 400–409. http://dx.doi.org/10.1016/j.neuropsychologia.2012.11.021.
Tononi, G., Sporns, O., & Edelman, G. M. (1992). Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. *Cerebral Cortex, 2*(4), 310–335. http://dx.doi.org/10.1093/cercor/2.4.310.
Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*, 171–178.
Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review, 95*(1), 15–48. http://dx.doi.org/10.1037/0033-295X.95.1.15.
Turnbull, C. M. (1961). Some observations regarding the experiences and behavior of the BaMbuti Pygmies. *The American Journal of Psychology, 74*, 304–308. http://dx.doi.org/10.2307/1419421.
Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: an object-superiority effect. *Science, 186*, 752–755. http://dx.doi.org/10.1126/science.186.4165.752.
Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*, 59–85. http://dx.doi.org/10.1016/0010-0285(70)90005-8.
Wolfe, J. M. (2012). The binding problem lives on: comment on Di Lollo. *Trends in Cognitive Sciences, 16*(6), 307–308.
Woodman, G. F., & Luck, S. J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychological Science, 14*(6), 605–611. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1472.x.

CHAPTER EIGHT

Technology-Based Support for Older Adult Communication in Safety-Critical Domains

Daniel Morrow[1]
Department of Educational Psychology, University of Illinois Urbana-Champaign, Champaign, IL, USA
[1]Corresponding author: E-mail: dgm@illinois.edu

Contents

Abstract

The study of how age-related changes in cognition influence communication and learning in safety-critical domains has important practical and theoretical implications. The work described in this chapter leveraged theories of communication and cognitive aging in order to investigate how collaboration between older adults and their health-care providers can be supported by technology. According to interactive theories of communication, collaboration requires providers and older adults to present, understand, and ground contributions so that these contributions are accepted as mutually understood and relevant to shared goals. These collaborative processes are influenced by resources and constraints associated with communication media (e.g., face-to-face versus asynchronous communication) and communication partners (e.g., health literacy abilities such as processing capacity and knowledge). We investigated how older adult/provider collaboration needed for self-care can be supported by technology. We built on benefits of face-to-face communication (e.g., rapid turn-taking, nonverbal cues to

Psychology of Learning and Motivation, Volume 64
ISSN 0079-7421
http://dx.doi.org/10.1016/bs.plm.2015.09.008

affective as well as cognitive meaning) while using technology to address constraints associated with this communication situation, such as transient speech and message reviewability.

1. INTRODUCTION

Age-related changes in cognition have many influences on communication and learning in safety-critical domains that have important implications for both practice and theory. The practical issues are pressing. First, our aging society is reflected in an aging workforce (http://www.bls.gov/spotlight/2008/older_workers/). For example, the average age of health-care providers (e.g., physicians, Young, Chaudhry, Rhyne, & Dugan, 2011) has been increasing, with those in their 60s the single most frequent age category. Our aging workforce poses many challenges to society. For example, it is important to sustain older workers' productivity at the individual (e.g., compensatory strategies that support individual performance), group (e.g., collaborative strategies that offset age-related cognitive declines among individuals), and organizational (e.g., management practices that increase worker flexibility) levels, so that older workers remain a source of organizational resilience (Morrow & Czaja, 2015). Our aging society also challenges the national health-care system. Age-related risk for chronic care in the context of a system that puts increasing responsibility on patients for their own care places heavy demands on older adults' ability to effectively manage their own illness. The retiring baby boomer generation in particular has many repercussions for the economy (increasing demands on social security and medicare; generational transfer of wealth).

A second major issue related to aging and communication is the rapid evolution of technology that has fundamentally reshaped society, including how information is consumed and shared. Technology may be even changing cognition, for the better (augmenting cognition by supporting the ability to retrieve, understand, and use information; providing training that can "exercise" and perhaps improve cognition), the worse (promoting distraction and shallow processing that undermines the ability to engage in sustained, critical thinking; Carr, 2011), or both. How can older adults engage new information ecologies in order to remain vital, fully participating members of society at work and home?

Finally, an aging society that must cope with rapidly evolving technology as well as structural changes that require switching careers across the life span, puts a premium on life-long and life-wide learning (National Research Council, 2012a). A barrier to such learning is the alarming estimates of inadequate adult functional literacy, especially for older adults (National Research Council, 2012b). In this chapter, I focus on how older adults can take advantage of technology to support learning and self-care.

Research that addresses these age-related issues also provides opportunities for advancing theory. In his influential book, *Pasteur's Quadrant*, Donald Stokes argued that addressing complex social problems requires "use-inspired basic research" (Stokes, 1997). Tackling issues such as improving productivity and self-care among older adults requires leveraging theories of how aging influences communication and cognition. At the same time, "use-inspired research" can spur theory development because the problems become a crucible for testing implications of these theories (Morrow & Durso, 2011). For example, theories of cognition have evolved from a focus on individual processes to considering how cognition is distributed across individuals and the environment in order to explain how individuals achieve high levels of performance in airplane cockpits, intensive care units, and other complex socio-technical settings. These theories analyze how performance emerges from the interaction of internal and external representations of information as people interact with complex systems (Coiera, 2006; Hazlehurst, Gorman, & McMullen, 2008; Hutchins, 1995). In cognitive aging theory, the concept of environmental support developed in order to explain how older adults sometimes perform complex memory-based tasks such as taking medication as well as younger adults do, despite age-related declines in the efficiency of basic encoding and retrieval processes (Craik, 1994; Morrow & Rogers, 2008).

In the rest of the chapter, theories related to communication and aging are overviewed in order to frame a summary of my research related to the effects of aging on communication in complex socio-technical domains. The focus is on health care, but comparisons are also made to aviation in order to generalize findings across domains when possible.

2. THEORIES OF COMMUNICATION

2.1 Communication Processes

Similar to the evolution from individual to distributed cognition theories, theories of communication in socio-technical systems have evolved from viewing communication as message transmission between individuals to viewing it as interactive, coordinated action. These two approaches view the cognitive processes and resources involved in communication in different ways. Derived from an information-theoretic analysis of communication (Shannon & Weaver, 1949), the information transmission view analyzes message production and comprehension as largely independent activities, with production limited by speaker cognitive resources (e.g., working memory constraints on encoding and sending messages) and comprehension limited by listener resources (e.g., working memory constraints on decoding messages; Wickens & Hollands, 1992). In the interactive view, communication partners coordinate resources in order to co-construct meaning, so that producing and understanding messages are intertwined processes (Clark, 1996; Pickering & Garrod, 2006). Speakers engage their addressee's attention with verbal or nonverbal (e.g., gestures) cues in order to present messages that are sensitive to the partners' common ground (shared knowledge of language, culture, the prior discourse, and communication context). Listeners not only understand the message, but signal that they do so, either implicitly by responding with a relevant contribution or explicitly by signaling that they understand (e.g., a head nod). They may also signal lack of comprehension either implicitly (e.g., a puzzled expression) or explicitly (e.g., by requesting clarification). Speakers and listeners accept the message as mutually understood and relevant to shared goals, thereby updating common ground (Clark, 1996). Processes involved in presenting, understanding, and accepting messages often overlap and interact. For example, a physician may immediately rephrase "your Lipitor" as "your cholesterol pills" in response to a patient's puzzled expression before completing her utterance.

Grounding, or "closing the communication loop," is essential for successful communication, whether it supports immediate action (when flying an aircraft or operating on a patient) or longer term goals such as learning in the classroom. Thus, communication depends on partners' joint effort to coordinate their contributions. Establishing common ground is essential for safety in complex settings. For example, when nurses change shift in the ICU, the outgoing nurses who hand off patients to incoming nurses may overestimate common ground if they know that the incoming nurses are familiar with the patients, so that they overly abbreviate the shift report. This strategy may minimize their own effort involved in presenting information during the hand off, but at the expense of the incoming nurse who is likely to misunderstand and to request clarification, which increases both partners' effort to reach common ground (Carroll, Williams, & Gallivan, 2012). Such communication problems in turn contribute to patient safety incidents (The Joint Commission, 2005). More generally, communication problems are reported as contributing factors in at least 70% of adverse events that threaten patient safety in hospitals and other health-care settings (Cordero, 2011). Similarly, in aviation, air traffic controllers, who direct multiple aircraft by issuing voice instructions to the aircraft on the same radio frequency, may attempt to minimize their own effort by producing one long message to an aircraft rather than dividing it into two or more shorter messages, which would require more radio time. However, this strategy increases the likelihood that the intended addressee will misunderstand this long message and have to request clarification, thus increasing the cognitive effort (and radio time) of both controller and pilot to repair communication (Morrow, Rodvold, & Lee, 1994). As in health care, these communication problems erode aviation safety (Cardosi, Falzarano, & Han, 1998). In short, successful communication depends on partners' ability to coordinate contributions. The literature on communication in complex settings suggests that communication efficiency and accuracy depends on resources afforded by different communication media situations, as well as by communication partner abilities.

2.2 Communication Media Resources

Communication media vary in the resources afforded to and constraints imposed on partners' attempts to reach common ground (Clark & Brennan, 1991; Monk, 2009). In face-to-face communication, the most basic situation, partners can see and hear each other as well as the communication context. They use nonverbal cues (gesture and facial expression) as well as verbal cues to coordinate attention when presenting, understanding, and accepting messages. Therefore, communication can be efficient, with less need for elaborate description compared to other communication situations (Gergle, Kraut, & Fussell, 2004). Grounding is also supported because partners' contributions are typically co-temporal and can be simultaneous, so that problems can be quickly uncovered and repaired. The intertwined nature of presenting, understanding, and grounding contributions in face-to-face communication is reflected in the fact that partners often align their choice of words, syntactic structures, and descriptions (Pickering & Garrod, 2006).

Of course, face-to-face communication also has drawbacks. Communication is often multifaceted, with partners needing to track multiple intertwined topics. It is also often interleaved with concurrent tasks such as driving. Partners can have competing as well as shared goals, especially in settings that involve interdisciplinary teams with members that have distinct as well as overlapping expertise, such as nurses and physicians from different specializations during surgery. The need to easily access concepts from prior discourse (reviewability) and revise contributions in light of the evolving discourse (revisability) is challenging because speech is transient, and the time pressure of face-to-face communication can hinder the deliberation needed to craft complex messages (Clark & Brennan, 1991). For these and other reasons, decision-making and problem solving can be worse in groups using face-to-face communication compared to nominal groups (the pooled performance of the same number of people working alone (Ekeocha & Brennan, 2008; Weldon & Bellinger, 1997).

Face-to-face communication in complex settings such as health care is more often the exception than the norm. Communication is often distributed over space (synchronous), supported by mobile phone, video conference, or chat, as well as over time (asynchronous), supported by technology such as e-mail and voice mail. Interactive theories are helpful for analyzing distributed as well as face-to-face communication (Clark & Brennan, 1991; Monk, 2009). Synchronous distributed, like face-to-face communication, allows rapid turn-taking, co-temporality, and simultaneity that support grounding, but it may eliminate visual cues (unless video conferencing), which can increase collaborative effort involved in accepting contributions (Clark & Brennan, 1991). For example, communication workload may increase when partners are not copresent because speakers cannot modulate their contributions as required by listener context, as seen in the finding that conversation is more distracting to drivers who talk to someone on a cell phone than in the car (Drews, Pasupathi, & Strayer, 2008). Texting and chat can support rapid turn-taking that approximates face-to-face communication, but they eliminate auditory cues (e.g., speech prosody) as well as visual copresence, which can complicate grounding messages. However, they do provide a message record that supports reviewability and revisability. An important goal of designing synchronous communication systems is to incorporate resources associated with face-to-face communication, such as partner copresence, although doing so poses technical and design challenges (Monk, 2009).

Grounding is even more challenging for asynchronous communication, which lacks co-temporality, simultaneity, and sometimes sequentiality of turns. More effort may be needed to produce messages (typing vs speaking), which influences individual and collaborative effort. Asynchronous voice communication (exchanging prerecorded messages) is often less effective than telephone or other synchronous communication. This situation especially challenges grounding to the extent that shared communication and task goals are not saliently represented in the technology that supports communication (Collins, Bakken, Vawdrey, Coiera, & Currie, 2011).

2.3 Communication Partner Resources

Presenting, understanding, and grounding messages depend on partners' cognitive abilities as well as communication media constraints. Presenting messages is constrained by limited processing capacity (e.g., working memory), both at the level of producing utterances and more globally, planning how the utterances fit into the unfolding conversation (e.g., Levelt, 1989). Processes involved in understanding these contributions (recognizing words, activating and integrating concepts into propositions, and elaborating these ideas with knowledge to create situation models) are similarly constrained by the listener's processing capacity (e.g., Kintsch, 1998). For example, patterns of pausing suggest speakers or writers plan their utterances around syntactic units such as clauses, reflecting working memory constraints. Similar patterns of pausing during reading suggest readers take time at the end of clauses in order to "wrap up" their understanding by integrating concepts and by drawing inferences to elaborate propositions into the situation model or a gist-based representation of the message (Stine-Morrow & Miller, 2009). While processing capacity limits constrain communication, knowledge often facilitates message production (Levelt, 1989) and comprehension (Kintsch, 1998) by making processes more efficient and less dependent on processing capacity.

Most important, communication partners coordinate their cognitive effort to accomplish joint goals, so that shared attention is a critical collaborative resource for accepting contributions into common ground and updating a shared situation model. Grounding processes vary in the extent to which they require cognitive resources. Collaboration can involve strategic negotiation of what is accepted as mutually understood, requiring resources to update common ground (Clark, 1996). Collaboration can also be a more implicit process accomplished with little effort, with default choices that align partners' representations until one partner recognizes that their representation is inconsistent with the other's, which triggers effortful repair (Pickering & Garrod, 2006). This may especially be the case for the situation model-level of understanding, which involves perceptual simulations of described situations (Zwaan & Taylor, 2006). Such relatively automatic alignment of partners' situation models may be likely for face—face communication, where verbal and nonverbal cues support monitoring and other grounding processes (Bavelas & Gerwing, 2011). It may be less likely for asynchronous communication unless copresence is promoted by teleconferencing technology (Monk, 2009).

2.4 Aging and Communication

Age-related changes in cognition influence communication. Declines in "mental mechanics" or processing capacity (e.g., working memory capacity) can impair the ability to interpret and reason about novel information (Baltes & Baltes, 1990). Age-related processing capacity constraints can undermine processes involved in presenting messages. There is some evidence that older adults (especially in their later 70s and older) produce syntactically simpler utterances and shorter discourse with less clearly identified relationships between ideas (e.g., ambiguous pronouns), although older adult communication performance is highly variable (for a review see Bloom, Obler, De Santi, & Ehrlich, 2013). Similarly, age-related declines in processing capacity impair comprehension, especially for syntactically complex and conceptually dense text that challenges conceptual integration (Stine-Morrow & Miller, 2009). In contrast to processing capacity, knowledge tends to increase with age-related experience related to literacy engagement and domain-specific learning, and can offset the effects of processing capacity limits on older adults' comprehension, especially at the situation model or gist-based level of comprehension (Payne, Gao, Noh, Anderson, & Stine-Morrow, 2012; Stine-Morrow & Miller, 2009). Older adults' comprehension of health information in particular is influenced by the interplay of age-related losses in processing capacity and gains in knowledge about language and health (Beier & Ackerman, 2005; Chin et al., 2011; Chin, Madison et al., 2015).

Age-related changes in cognition influence grounding as well. Limited processing capacity may constrain older adults' ability to present messages that are sensitive to common ground. They may not shorten referring expressions as efficiently as younger adults do during conversation, perhaps reflecting age-related difficulty retrieving information about their partner and the prior discourse context (Horton & Spieler, 2007; Hupet, Chantraine, & Nef, 1993). However, age differences in the efficiency of grounding may be overestimated in these studies because of abstract tasks such as the communication matching game, where one partner directs the other to arrange a set of abstract objects so that the addressee's order of objects matches their own order, but the partners are not copresent in the situation. Age differences in grounding processes may be reduced when partners share knowledge about the topic and context. For example, "collaborative success" on recall and problem solving tasks (better collaborative group performance compared to nominal group performance) occurs for experts but not for nonexperts among young adults, in part because expert partners jointly elaborate contributions based on their shared domain knowledge (Meade, Nokes, & Morrow, 2009). Older adults can benefit at least as much as younger adults do from collaboration (Henkel & Rajaram, 2011), although expertise-related benefits for older adults have not been investigated. In addition, age differences in story recall were eliminated in one study when participants retold the story to a child rather than the experimenter, suggesting that older adults maintain the ability to monitor the communication situation and design contributions relevant to their audience, at least for face-to-face communication (Adams, Smith, Pasupathi, & Vitolo, 2002).

In addition to shared knowledge, another powerful strategy for supporting older adults' communication is off-loading cognitive workload to the environment, as suggested by theories of distributed cognition and environmental support (Morrow & Rogers, 2008). Relying on external representations may minimize cognitive constraints on communication in several ways.

1. External representations provide rapid access to information from perception rather than from memory (Gray & Fu, 2004). Perceptual access can facilitate communication: a shared visual context during conversation reduces need for elaborate verbal descriptions because speakers can simply point at referents (Gergle et al., 2004). External representations should especially help older adults to ground messages to the extent processes such as designing messages for addressees are limited by inefficient memory retrieval (Horton & Spieler, 2007). Perceptual access also supports older adults' comprehension. For example, adults with hippocampal damage that impair declarative memory are able to collaborate as well as matched normal adults do by quickly identifying the speaker's intended referent, as long as the referent information is readily accessible from perception rather than memory (Rubin, Brown-Schmidt, Duff, Tranel, & Cohen, 2011). This may also be the case for normally aging adults because aging is associated with impaired hippocampal function (Daselaar, Fleck, Dobbins, Madden, & Cabeza, 2006). More generally, shared visual context supports joint focus of attention, so that grounding is more implicit and less dependent on processing capacity, and partners'

alignment of linguistic and situational representations is relatively automatic (Pickering & Garrod, 2006).

2. External representations can also reduce communication demands on processing capacity to the extent that information is perceptually available and therefore does not need to be inferred or computed. For example, text is more accurately understood when accompanied by a graphic that explicitly organizes concepts mentioned in the text, which reduces the conceptual integration and elaboration processes needed to develop the situation model (Glenberg & Langston, 1992; Larkin & Simon, 1987).
3. Shared external representations also help partners align their perspectives on a task, which can reduce discrepancies between their situation models needed to accomplish the task (Ertl, Kopp, & Mandl, 2008; Fischer, Bruhn, Gräsel, & Mandl, 2002). This may especially be the case when the representations are co-constructed during collaboration (Schwartz, 1995). Shared external representations may also encourage use of knowledge to accomplish joint goals by helping to activate shared knowledge relevant to the task. For example, the use of graphic scripts during problem solving increases knowledge convergence among partners during collaboration (Ertl et al., 2008).
4. External representations can reduce the costs of interruption during communication because partners find it easier to switch back from the interrupting task to the ongoing communication task if information is externally represented, so that resuming communication involves reaccessing information from perception rather than from memory. For example, checklists help people recover from interruptions when performing complex tasks (e.g., operating on patients, flying an aircraft; Gawande, 2010).

In summary, while age-related declines in processing capacity can impair the ability to present, understand, and ground contributions to collaborative tasks, age-related differences may be mitigated when shared knowledge and external representations offset processing capacity constraints. Benefits of external representations for older adults may be more likely for face-to-face than for distributed communication because of opportunity for direct, perception-based interaction with these representations to support collaboration, although well-designed technology can afford similar opportunity for collaboration in synchronous, and perhaps asynchronous, distributed communication situations by promoting partner copresence (Monk, 2009).

3. STUDIES OF COMMUNICATION BETWEEN HEALTH-CARE PROVIDERS AND OLDER ADULTS

I next summarize several studies that investigate the use of technology to support older adult/provider collaboration required to self-manage chronic illness such as hypertension and diabetes, focusing on older adults' ability to present, understand, and ground information about self-care tasks. Designing technology to support collaboration requires analysis of the target collaborative tasks, and how the processes in these tasks depend on resources related to the communication media and partners' cognitive abilities. Studies that focus on technology supporting face-to-face communication between providers and older adults are described first, followed by studies of distributed communication.

3.1 Supporting Face-to-Face Communication

Older adults often take multiple medications to treat symptoms and slow illness progression of chronic illness. For example, older adults with type II diabetes take four to five daily medications (Budnitz & Layde, 2007). More complex medication regimens (more medications and daily doses) are associated with lower adherence among older adults (Ingersoll & Cohen, 2008), in part because of a mismatch between the cognitive demands of self-managing multiple medications and the literacy and cognitive resources that older adults bring to this task (Morrow & Wilson, 2010). Adherence requires patients not only to understand how to take each medication, but to organize this knowledge into a plan for taking the medications together (Morrow & Wilson, 2010; Wolf et al., 2011). Implementing the plan can also be demanding: Patients must encode intentions to take each medication, act on these intentions at appropriate times, and monitor to ensure the medications were actually taken (Insel, Morrow, Brewer, & Figuerosa, 2006).

Ideally, providers collaborate with patients to help them develop adherence plans. Medication review is a key to managing older adults' complex regimens (Aspden, Wolcott, Bootman, & Croenwett, 2007). This requires provider and patient to review each medication in the patient's regimen to ensure that all medications are current and appropriately prescribed, resulting in an up to date and comprehensive medication list. This "medication reconciliation" process is especially important during care transitions when medications can change, such as returning home from the hospital (Aspden et al., 2007). Collaborative medication review is also an important opportunity to educate patients by checking that they understand how to take the medications safely (e.g., avoiding drug interactions or conflicts with other activities such as meals). Unfortunately, face-to-face patient/provider collaboration is often inadequate because of limited provider time and inadequate training of communication skills such as considering patient goals and perspective and working with patients to develop self-care plans (Haidet, 2007). Presentation of key information is often incomplete (Tarn et al., 2006) and information may not be presented in a patient-centered manner (use of jargon, Stewart, 1995). Grounding of this information is also inadequate: Physicians rarely check that their patients understand presented information (Schillinger et al., 2003; Schwartzberg, Cowett, VanGeest, & Wolf, 2007). Medication reviews in particular are often sporadic, incomplete, and poorly organized, which contributes to medication errors and reduces outcomes (Tarn et al., 2006; Tarn, Paterniti, Kravitz, Fein, & Wenger, 2009).

According to common ground theory, effective medication review requires provider and patient to develop a shared plan for taking medications. Developing this shared plan requires the provider to present the information that the patient needs to take their medications safely and the patient to talk about possible barriers (concern about side effects, conflicts with daily routine). Providers and patients need to work together to integrate the medication and patient information into a plan (e.g. which medications to take together, with or without meals or other daily events), and then evaluate, revise, and accept the plan as mutually understood. Developing the plan involves more local contributions such as presenting, understanding, and grounding medication facts (e.g., how many times a day to take; how many pills). Older adults may have trouble understanding (e.g., activating and integrating the concepts that constitute these facts) and integrating the information into a plan because of processing capacity limits, especially when providers use conceptually dense language with jargon (Chin, Payne et al., 2015). While prior knowledge (about language and health topics) can facilitate comprehension and potentially mitigate the impact of processing capacity constraints on comprehension (Chin, Madison et al., 2015; Chin, Payne et al., 2015), older adults with low health literacy have limited knowledge as well as processing capacity. Moreover, providers do not always have time to effectively exploit the resources associated with face-to-face communication (e.g., monitoring patient cues to ground information), and do not address inherent constraints of face-to-face communication such as limited message reviewability in order to support collaboration.

Guided by distributed cognition and cognitive aging theories, my colleagues and I developed a tool (the MedTable) to help providers and older adults collaborate during face-to-face encounters in order to develop, evaluate, and revise medication adherence plans (see Figure 1). Most generally, the tool provides a shared visual display that addresses the reviewability and revisability constraints associated with face-to-face communication (Clark & Brennan, 1991). A table format is used because this format can be effective for representing scheduling task constraints (e.g., Day, 1988). Each row summarizes key medication information (e.g., number of times a day to take) and each column represents a time or event from the patient's daily routine (e.g., wake up, meal time). The table should support the ability of provider and patient to present and understand information needed for developing the plan by cueing memory for key medication and daily routine information. More generally, graphics can support problem solving because they can be more computationally efficient than text, reducing need to search for and integrate information. For example, if relationships between medication constraints (take twice a day with food) and daily routine constraints (meal times) are directly represented in the table, there is less need to keep constraints in working memory when developing the plan (Larkin & Simon, 1987; Morrow, Hier et al., 1998).

The table also supports collaborative processes by providing a shared workspace that helps patient and provider coordinate attention when developing the plan. Patient and provider can simply point to rather than describe information on the table (e.g., previously scheduled medications, daily events). Because providers see how patients plan to take their medications, the tool supports strategies that support grounding, such as teach back and teach-to-goal strategies, in which patients demonstrate (verbally or nonverbally) their understanding so that providers can check their comprehension (Schwartzberg et al., 2007). Jointly completing the table encourages patient and provider to review and revise medication schedules. For example, more optimal schedules (consolidated around fewer medication times with equal spacing between doses) may look simpler in the table. This tool may especially benefit older adults because, without help, they tend to create overly complex schedules (Wolf et al., 2011). Moreover, anchoring schedules to daily routines helps patients think concretely about how to take their medications, which can improve prospective memory (Liu & Park, 2004).

To evaluate the MedTable, we started "low tech" with a paper version of the tool in a lab-based collaborative medication planning task (Figure 1; Morrow, Raquel et al., 2008). Older adults were randomly assigned to work

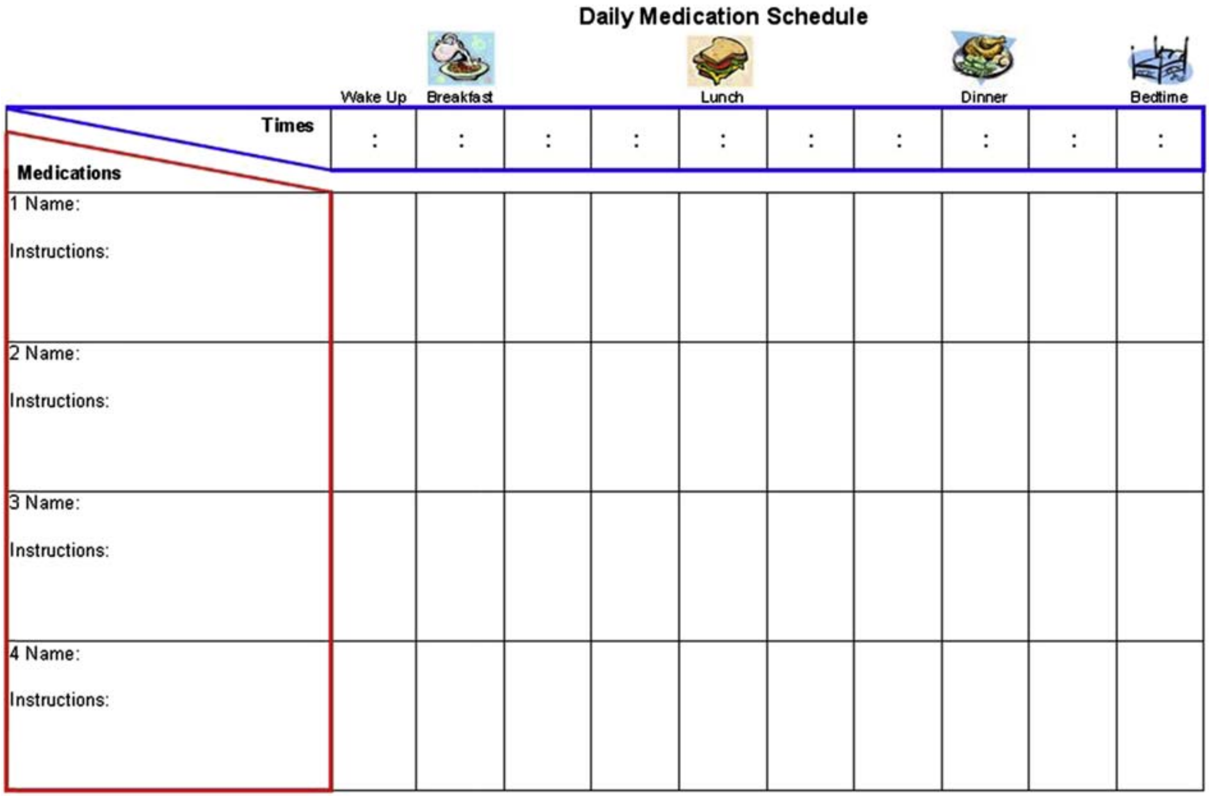

Figure 1 Paper prototype for MedTable tool. *From Morrow, Raquel, et al. (2008). Reprinted with permission.*

in pairs, one as "provider" and one as "patient." The providers received information about the medications to be scheduled (e.g., number of times to take per day) and patients received information about their daily routine (e.g., when they wake, eat meals, work). The pair worked together to create a schedule that met the constraints, either using the MedTable, an unstructured tool (blank paper), or no tool (talk only). They created more accurate schedules when using the MedTable, presumably because it reduced the cognitive demands of creating the schedules. Analysis of participants' conversation showed that, compared to the blank paper condition, the tool supported more structured and interactive conversation (less information needed to be verbally presented and there was more explicit confirmation of presented information), and review of proposed schedules was more frequent. In a follow-up study, actual providers (physician, nurse, or pharmacist) and patients used the tool to review the patient's regimens. Most patients were able to organize their regimens around their daily routine using the tool, and thought the tool was helpful and easy to use (Conner-Garcia, Morrow, Graumlich, Ellison, & Wang, 2012).

The MedTable needs to be flexible in actual practice so that providers can create schedules for patients with different medication regimens and daily routines, and update the MedTable-based schedules as regimens change over time. We developed an electronic MedTable™ that replicated key features of the paper tool (e.g., table format representing medication and patient routine constraints) and allowed providers to input and store patient-specific medication and daily routine information, and click on table cells in order to create the schedule (Kannampallil, Waicekauskas, Morrow, Kopren, & Fu, 2013). Using the lab-based collaborative medication scheduling task, we compared the computer-based and paper-based MedTables to a less structured paper-based tool: a card that presented medication information in a list rather than table format. Such cards are often used for medication review and reconciliation during patient visits (Aspden et al., 2007), although the list format may not be optimal for schedule tasks (Day, 1988). Participants in the study created accurate schedules regardless of which tool they used, but the two structured tools were easier to use (lower workload and higher usability ratings). Moreover, the schedules created using the two structured tools were simpler, with medications taken at fewer daily times and with more evenly spaced doses (Kannampallil et al., 2013).

Finally, the computer-based MedTable™ was integrated into an electronic medical record (EMR) system in order to facilitate use in clinical contexts (Figure 2; Morrow et al., 2012). Implementing the tool required

MedTable™
Northwestern Medical Faculty Foundation

361 05/31/2011

Medicines	6:30 AM	7:00 AM	7:30 AM	8:00 AM	8:30 AM	9:00 AM	9:30 AM	10:00 AM	10:30 AM	11:00 AM	11:30 AM	Noon	12:30 PM	1:00 PM	1:30 PM	2:00 PM	2:30 PM	3:00 PM	3:30 PM	4:00 PM	4:30 PM	5:00 PM	5:30 PM	6:00 PM	6:30 PM	7:00 PM	7:30 PM	8:00 PM	8:30 PM	9:00 PM	9:30 PM	10:00 PM	10:30 PM	11:00 PM	11:30 PM	Midnight	24:30	25:00
INSULIN ISOPHANE HUMAN 100 UNIT/ML SUSP controls blood glucose			100																																			
GLYBURIDE 5 MG TABS lowers blood glucose			1																																			
OMEPRAZOLE 20 MG CPDR •treats ulcers •treats Gastroesophageal Reflux Disease (GERD)		1																							1													
CYCLOBENZAPRINE HCL 10 MG TABS •is a muscle relaxant •relieves pain			1																							1												
NAPROXEN 375 MG TBEC •relieves pain, swelling and stiffness •treats osteoarthritis and rheumatoid arthritis			1																							1												
DICYCLOMINE HCL 10 MG CAPS treats irritalble bowel syndrome			1																							1												
MELOXICAM 7.5 MG TABS treats osteoarthritis and rheumatoid arthritis			1																																			
DULOXETINE HCL 60 MG CPEP treats depression and anxiety			1																																			
FERROUS SULFATE 325 (65 FE) MG TABS			1																																			
AMITRIPTYLINE HCL 50 MG TABS treats depression			1																																			
RANITIDINE HCL 300 MG TABS treats ulcers and gastroesophageal reflux disease (GERD)																											1											

Figure 2 Electronic medical record-based electronic version of the MedTable tool. *From Morrow et al. (2012). Reprinted with permission.*

addressing system barriers to medication reconciliation such as variable and technical language used in patients' medication lists stored in the EMR. A protocol was developed to use the tool to address these barriers and support collaborative planning. Before the patient's visit to the clinic, a nurse loaded the patient's EMR medication list into the tool, simplifying the technical information for patients with low health literacy. During the visit, the nurse and patient reviewed the patient's medication list and updated it as needed (reconciliation). Next, patient and nurse used the tool to jointly review the patient's current medication schedule. To do this, the nurse set up the MedTable around the patients' daily routine, with key daily events (e.g., meals) represented by icons. Next, the nurse and patient discussed and scheduled each medication by clicking on the cell corresponding to the medication (row) and time slot (column) for the chosen time. The relevant rows and columns were highlighted so that the results of user actions were easily visible. Therefore, presentation of key information by both provider and patient was supported. Most important, the tool was designed to encourage both partners to evaluate and revise the proposed schedule, supporting grounding of the information. For example, the nurse could suggest more effective schedules (e.g., all medications to be taken with food were presented together in the table so they could be scheduled at the same time and linked to an event). Because the agreed-upon schedule was readily visible to both provider and patient, the display encouraged use of teach back and teach-to-goal strategies to improve patient comprehension and reinforce grounding of the plan. At the end of the visit, the patient took a printed copy of the schedule home to use as a guide to adherence.

A randomized trial was conducted to evaluate the impact of the MedTable-based intervention relative to a usual care control group on diabetic patients' knowledge about their medications, self-reported medication adherence, satisfaction with provider communication about medications, and HbA1c levels (Morrow et al., 2012). Data collection is complete and currently being analyzed. If the intervention proves successful, an important next step would be to integrate the MedTable system with patients' mobile phones, perhaps in the form of a medication schedule app (cf. Siek et al., 2011). A phone-based system would address several problems with the current MedTable implementation. First, the tool is implemented on a desktop computer, which does not easily support joint use of the tool by provider and patient (compared to other technology such as a large touch screen). Second, an app linked to an EMR would support ongoing patient/provider collaboration through asynchronous communication when the patient is at home, especially if the system is integrated with the patient's personal health record. This would allow patient and provider to jointly update the patient's medication schedule as the patient's regimen changes over time, which often happens for older adults with chronic illness treated by multiple providers. Third, an app could be programmed to remind patients when to take their medications, thus supporting their ability to implement as well as revise plans. Of course, this type of app would need to be carefully designed to support the specific collaborative processes described above (e.g., presenting medication information, developing, revising, and accepting schedule changes) in distributed as well as face-to-face communication situations.

3.2 Supporting Distributed Communication

Delivery of health-care services is increasingly distributed, with patients remotely accessing and receiving services, in order to meet the needs of a mobile society and to address the problem of dwindling patient/provider contact time. Distributed communication is enabled by rapidly developing technology that supports both synchronous (telephone, video conferencing; texting, chat) and asynchronous (e.g., voice mail, e-mail, Web-based messaging integrated with EMR systems) communication between patients and providers. Distributed communication has the potential to transform health care by supporting patient-centered and collaborative care (IOM, 2012; Stead & Linn, 2009). For example, the medical home model of care delivery, in which patients receive coordinated remote care from multiple providers at home, is enabled by EMR systems (Baek & Seidman, 2015). However, this potential has yet to fully materialize. Older adults have been especially slow to adopt new technology-based services such as personal health records and Web-based patient portals to EMRs, in part because of cognitive and health literacy-related limitations (for review, see Morrow & Chin, 2012). My colleagues and I have conducted several studies that address barriers to distributed communication between providers and older adults that relate to media constraints (e.g., noninteractive communication that challenges grounding) and communication partners' cognitive resources (e.g., demands of speech comprehension on working memory and meta-cognitive processes such as comprehension monitoring).

3.2.1 Automated Telephone Messaging

Telephone communication has supported asynchronous as well as synchronous distributed communication in primary care settings for many years. In the 1990s, automated telephone messaging with interactive voice response

(IVR) and voice mail technology proliferated as a way to support continuity of care in between patients' visits to their providers. Automated messaging is now used to remind (attend appointments, take medication), survey, and educate patients, and even to help manage chronic illness (Friedman et al., 1998; Piette et al., 2013; Whitten, 2001). Telephone messaging helps providers to present and patients to understand health information because the systems can be programmed to deliver complete, standardized information that can also be tailored to different types of patients. Reviewability and revisability constraints are addressed because voice messages are stored and repeated. Nonetheless, older adults may have difficulty understanding or interacting with voice messages because of the transient speech medium, potential for poor voice quality, and background noise in the home environment, coupled with age-related sensory and processing capacity declines (Morrow & Leirer, 2001). Moreover, although IVR features such as keypad responses to questions can support message acknowledgments and other grounding processes, there is a concern that the technology is more likely to support presentation than grounding of health information, so that information is pushed at patients without ensuring that the information is mutually understood and acted on (Morrow & Fischer, 2013).

We have investigated age-related differences in understanding, remembering, and responding to automated telephone messages about health services. Older adults listened to and answered questions about prerecorded voice reminders to attend health-care appointments or to take medication. The messages varied in length and organization. In well-organized messages, the information (e.g., appointment time, location, preappointment procedures) was presented in an order that matched participants' preferences for organizing appointment or medication information ("schema-compatible" organization). Older as well as younger adults better remembered shorter and schema-compatible messages, presumably because these messages reduced demands on processing capacity (e.g., working memory resources needed to integrate propositions into a text-based representation of the message; Morrow, Carver, Leirer, & Tanke, 2000; Morrow, Leirer et al., 1998). Moreover, the benefits of message organization were greater for questions requiring inferences than for questions about explicit information in the messages, which suggest patients could more easily integrate information into a situation model of the health-care task when the information was organized in terms of their prior knowledge. Older and younger adults in another study better remembered messages that were repeated, and differences in memory associated with age and working memory were smaller for the repeated messages, suggesting that repetition reduced demands on processing capacity (Morrow, Leirer, Carver, Tanke, & McNally, 1999a).

However, voice messaging systems in actual health-care systems typically involve optional repetition, so that patients only replay messages if they want to. More generally, patients often need to self-initiate repetition, note-taking, or other forms of environmental support when interacting with automated telephone or other systems in order to improve message comprehension. Older adults may not take full advantage of these opportunities afforded by interactive technology because of age-related declines in comprehension monitoring or other meta-cognitive processes (Dunlosky & Connor, 1997). Therefore, we examined possible benefits of optional message repetition and note-taking on memory for voice messages (Morrow, Leirer, Carver, Tanke, & McNally, 1999b). Both older and younger adults took advantage of optional repetition to improve their memory for the messages, but age differences in memory were not reduced by repetition. Similarly, both older and younger adults better remembered messages when they were able to take notes while listening to the messages (message memory was tested with notes present), but age differences were not reduced by note-taking.

These findings are consistent with the prediction that external support reduces demands of comprehension on processing capacity, but it is unclear why older adults did not take differential advantage of this environmental support to compensate for age-related differences in processing capacity in this study. A clue comes from work on the benefits of note-taking for pilots' memory for air traffic control information (e.g., instructions to change heading and altitude; Morrow et al., 2003; Morrow, Wickens et al., 2008). In these studies, pilots and nonpilots listened to voice messages from air traffic controllers that were delivered over radio. They "read back" or repeated each message so that the controller could verify comprehension (a standard procedure designed to ensure grounding of critical navigation information during flight). Opportunity to take notes while listening to the messages was manipulated. Note-taking on a knee pad during flight is a standard communication procedure for pilots. Without opportunity to take notes in the study, age differences in read back accuracy were equivalent for pilots and nonpilots (although pilots were more accurate overall) and were largely explained by processing capacity measures (e.g., working memory). However, with note-taking, age differences were eliminated for pilots but not for nonpilots. Pilots' notes were more complete and better organized than nonpilots' notes, and aging only influenced the nonpilots' notes.

Moreover, note-taking reduced the impact of processing capacity constraints on read back accuracy. This suggests that with experience, older adults can become at least as effective as younger adults at taking advantage of environmental support in order to offset the impact of processing capacity limits on performance. It is possible that training in using notes or other supports would increase benefits in everyday tasks, just as shared knowledge or expertise may increase benefits of collaboration for accomplishing tasks for young adults, and perhaps for older adults as well (Meade et al., 2009). When note-taking was integrated into cockpit computer technology in the form of an "e-pad" interactive display positioned near flight displays, it was used by older as well as younger pilots during simulated flight and age differences in read back accuracy were reduced compared to when participants had to rely on memory for the navigation information (Morrow, Wickens et al., 2008). Of course, another strategy to support older adults' comprehension in distributed communication situations is to shift from transient voice to more permanent visual media in order to address working memory limitations associated with comprehension. The final project described in this chapter explores a Web-based approach to support provider/older adult collaboration in distributed situations.

3.2.2 Patient Portals to EMRs

Health services are increasingly delivered via Web-based applications such as patient portals to EMRs and personal health records, which like automated voice messaging helps support patients' continuity of care in between provider visits (IOM, 2012; Stead & Linn, 2009; Tang, Ash, Bates, Overhage, & Sands, 2006). In these systems, Web-based messaging and e-mail link providers and patients (Morrow & Chin, 2012; Tang et al., 2006). Like automated telephone messaging, EMR portals increase access to health information (appointment notices, test results, medication lists). However, portals often serve more as repositories of information than as support for provider/patient collaboration. This may be one reason why portals are underutilized by older adults (Morrow & Chin, 2012). Older adults are not only less likely to use portals, they have trouble finding or understanding information when they do access their portal. For example, quantitative information such as test results is often provided through portals with limited context to support interpretation (e.g., a table of numbers). Patients with limited numeracy skills struggle to understand and reason about such numeric information (Peters et al., 2009; Taha, Czaja, Sharit, & Morrow, 2013).

Traditionally, patients with inadequate health numeracy and literacy turn to their providers for help interpreting numeric information such as test results. During face-to-face discussion, providers can use both verbal commentary and nonverbal cues (facial expressions, tone of voice, gesture) to convey a bottom-line interpretation of risk associated with the test results. For example, do a patient's HDL and LDL cholesterol scores suggest that their risk for cardiovascular illness is increasing or decreasing over time? In this way, providers help patients develop a gist-based understanding of their results. Gist-based representations are more qualitative than verbatim representations of numeric information (is risk high or low; has it increased or decreased?) and often incorporate affective and evaluative dimensions (Reyna, 2011). Gist-based understanding of health information, supported by provider—patient communication, may also mesh with patients' beliefs about their illness and its impact on their life (causes, consequences) to develop more adaptive illness representations (Morrow & Chin, 2015). Indeed, provider nonverbal behaviors during patient consultations are associated with patient satisfaction and outcomes (Ambady et al., 2002). Unfortunately, providers have less and less time to help their patients understand critical health information. In addition, while health technology such as portals is intended to support patient self-care, it may exacerbate rather than compensate for the effects of dwindling face-to-face communication because portal-based health information is often stripped of the clinical context, so that patients are confronted with numbers without the benefit of face-to-face communication with their providers to help them interpret the numbers.

We are investigating how to help older adults' understand portal-based numeric health information. Our general approach is to improve patient gist comprehension in asynchronous communication between providers and patients by taking advantage of collaborative resources available in face-to-face communication. We are leveraging progress in using computer-based agents (CAs) to promote learning in different contexts. CAs are onscreen characters with human-like behaviors such as talking and gesturing. They have been found to improve student learning in cognitive tutoring settings (e.g., Mayer & DaPra, 2012). They can also improve patient understanding of and adherence to self-care information (e.g., Bickmore et al., 2010). Members of our team previously developed a CA that uses nonverbal cues (voice stress and intonation; facial expressions) to convey affective as well as cognitive meaning (Huang, Hasegawa-Johnson, Chu, Zeng, & Tang, 2009). We are refining this CA to serve as a portal-based clinical intermediary that provides succinct commentary about risk associated with test results (or other patient information). The CA is designed to enhance, rather than replace, patient—provider collaboration. It provides a multimedia format with verbal and nonverbal cues that reinforce each other so as to offset the impact of age-related sensory and cognitive declines on patient comprehension and decision-making (Van Gerven, Pass, Van Merriënboer, Hendriks, & Schmidt, 2003). Nonverbal cuing of affective information may especially support older adults' intention to act because they often rely on affect when making health decisions (Peters et al., 2009). CAs are also more likely than text-based interfaces to engender social responses such as trust to automated systems (Pak, Fink, Price, Bass, & Sturre, 2012). CA-based messages, like automated telephone messages, can also be repeated in order to address reviewability and revisability constraints that can limit face-to-face communication. Thus, CAs can combine face-to-face and distributed communication resources to support collaboration.

To develop the CA for patient portals, we first video recorded a physician who described the results of cholesterol and diabetes screening tests for several patients. Because the video physician served as a template for refining the CA, we conducted a pilot study to evaluate whether the recorded physician's messages were understood and perceived as helpful by older adults (Azevedo et al., 2015). Older adults in the pilot study viewed the recordings of the physician describing the results of cholesterol and diabetes tests that conveyed low, borderline, or high levels of risk for illness. Because risk levels associated with test results depend on patient characteristics (e.g., age, gender, history of smoking, family history of heart disease), each message was preceded by a fictitious patient profile followed by the results of either a cholesterol or diabetes screening test for that patient. Participants read each profile and then viewed the video message about that patient's results. They then identified the level of risk described by the physician and indicated reactions to the messages' content and presentation. Participants thought that the physician's delivery of the message (facial expressions and tone of voice) was appropriate for the message content (1—9 scale where 9 is very appropriate: mean of 8.5 for cholesterol and 8.6 for diabetes messages). They also reported feeling stronger positive emotions after listening to the lower risk messages, and stronger negative emotions after the higher risk messages, with intermediate emotions reported for the borderline risk messages. These findings suggest that participants were able to create a gist-based representation organized in part around evaluative affect, based on the physician's nonverbal cues. Participants also accurately understood the gist for overall risk from the messages (lower, borderline, higher), and thought the messages were useful (Azevedo et al., 2015).

We are now comparing older adults' comprehension of and reactions to clinical test result messages presented in a standard portal format (table of numbers with minimal information about the scale of the numbers) to verbally enhanced, graphically enhanced, and video-enhanced formats. The verbally enhanced format includes labels for evaluative categories (more or less risk) that provide context for interpreting the numbers and promoting affective processing of the quantitative information (Peters et al., 2009). In the graphically enhanced condition, graphics that convey key relational features (larger/smaller than) that support gist understanding of risk are included (Garcia-Retamero & Galesic, 2009). Finally, in the video-enhanced condition, the same graphics are accompanied by the recorded physician providing commentary about the results, with nonverbal cues (prosody, facial expressions) signaling information relevance and guiding affective interpretation, as in ideal face-to-face communication. The graphics and numbers are also included in this condition, with visual cues that link the video commentary with relevant information in the graphic. In the current experiment, the video-recorded physician serves as proxy for the CA. A later experiment will compare the video to the CA developed from the video physician.

Findings from this study should inform strategies for helping older adults understand key self-care information provided in asynchronous communication settings such as EMR portals. Because the CA is generative, it will be more flexible and robust than video for supporting patient—provider collaboration. For example, physicians will be able to program the CA to discuss and explain a wide range of health information provided to their patients through the portal. An important next step would be to make the CA interactive (see Bickmore et al., 2010) so that technologies that replicate other face-to-face resources (e.g., turn-taking) can support grounding as well as comprehension of information, which is likely to increase patient engagement with and use of the information to support self-care.

4. CONCLUSIONS

Following a "use-inspired basic research" approach (Stokes, 1997), the work described in this chapter leveraged theories of communication and cognitive aging in order to address health-care challenges related to collaboration between older adults and their health-care providers. According to interactive theories of communication, collaboration requires providers and older adults to present, understand, and ground contributions so that they are accepted as mutually understood and relevant to shared goals (Clark, 1996). These processes are influenced by resources and constraints associated with the communication media (e.g., face-to-face versus asynchronous communication; Clark & Brennan, 1991) and communication partners (e.g., health literacy abilities such as processing capacity and knowledge; Chin et al., 2011). My colleagues and I have conducted research examining how to support older adult—provider collaboration needed for self-care by leveraging communication media and partner resources. We built on benefits of face-to-face communication (e.g., rapid turn-taking, nonverbal cues to affective as well as cognitive meaning) while using technology to address constraints associated with this communication situation, such as transient speech and message reviewability. The MedTable provides a shared visual workspace that supports the ability to present, understand, and ground the information needed to develop a shared plan for taking multiple medications. It is designed to support collaboration by being used by both provider and patient, rather than undermining collaboration by requiring providers' attention at the expense of communication with the patient (White & Danis, 2013). We are now testing whether an intervention based on this tool improves patients' medication knowledge and adherence.

Synchronous distributed communication supported by telephone or texting is often required in health-care systems, but can challenge collaboration. Older adults and providers are not copresent in this communication situation, which can exacerbate communication demands on partners' processing capacity that is imposed by transient voice. Our work on age differences in pilots' comprehension of radio-based messages suggests that experience using environmental support (such as note-taking) may mitigate demands on processing capacity during synchronous communication, although research is needed to examine this claim in health-care communication. Comprehension demands may also be mitigated by teleconferencing or other tools that support copresence in synchronous distributed communication, if the tools are designed to target specific collaborative processes. For example, performance in these situations is improved when the video conference link includes the communication context, so that partners not only see each other, but what they are looking at in the shared context (Monk, 2009).

Asynchronous communication may be most challenging for patient–provider collaboration, especially the ability to mutually accept information as understood and relevant so that older adults act on this information to support self-care. A danger of asynchronous communication is that technology such as patient portals or health Web sites distributes or broadcasts health information without ensuring that it is understood (Morrow & Chin, 2012). To help address this challenge, we used technology that enables communication resources that may reduce demands on older adults' processing capacity and leverage their knowledge. First, guided by theories of language comprehension, we supported older adults' use of automated telephone messaging systems. Older adults better remembered health-care appointment and medication reminder messages that were brief but still contained key information needed to accomplish the task, and that were organized to match knowledge related to the communication (schemas related to attending appointments and taking medication). We also integrated some features of face-to-face communication by making the systems more interactive, allowing people to repeat the messages. Repetition improved memory for messages and offset the effects of age-related declines in processing capacity on memory, although these benefits were attenuated when repetition was optional, perhaps because of age-related limits in comprehension monitoring or other meta-cognitive processes. More research is needed to determine how to design technology that supports meta-cognitive processes during asynchronous communication, such as embedding queries or other interactive features of face-to-face communication.

Another approach to support asynchronous communication between health-care providers and older adults is to more directly leverage resources from face-to-face communication, such as nonverbal cues that convey affective as well as cognitive meaning. Such cues may engage patients so that the patients act on as well as understand information needed for self-care. We are developing a CA in patient portals that emulates face-to-face communication to not only improve older adult comprehension of numeric health information (test results), but support collaboration with providers.

While these studies are guided by theories of how aging influences cognition and communication, they also raise challenges that may help spur theory development. First, to what extent can resources that support grounding in face-to-face communication (e.g., rapid turn-taking; nonverbal cues) be leveraged to support grounding in asynchronous communication? To the extent that these resources are basic to our interaction with the world, as suggested by theories of embodied cognition

(Barsalou, 2008), older adults may take advantage of them as effectively as younger adults do because the resources require little cognitive effort (Pickering & Garrod, 2006; Radvansky & Dijkstra, 2007). The goal is to leverage communication technology to create situations in which the interactivity of face-to-face communication is combined with benefits of asynchronous communication, such as support for message reviewability and revisability. This requires elaborating theories of communication in terms of theories of cognitive aging. Second, how are meta-cognitive abilities that are crucial for learning (Dunlosky & Connor, 1997) deployed in different technological environments to support learning and how can features of these environments be designed to support more effective monitoring? For example, the ability to monitor and control learning processes (e.g., restudy) may be less effective when studying online versus on paper (Ackerman & Lauterman, 2012). This problem may magnify age-related difficulties related to meta-cognitive processes involved in learning health information. Designing technology that supports rather than impairs the metacognition processes underlying learning requires combining the precision of process theories with the breadth of distributed cognition theories that identify key functional features of the technology that can support these processes.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by the National Institute of Aging (Grant R01 AG31718), National Institute of Nursing Research (Grant R01 NR011300), and the Agency for Health Care Research and Quality (Grant: R21HS022948). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NIH.

REFERENCES

Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure. *Computers in Human Behavior, 28*, 1816–1828.

Adams, C., Smith, M. C., Pasupathi, M., & Vitolo, L. (2002). Social context effects on story recall in older and younger women does the listener make a difference? *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57*, P28–P40.

Ambady, N., Koo, J., Rosenthal, R., & Winograd, C. (2002). Physical therapists' nonverbal communication predicts geriatric patients' health outcomes. *Psychology and Aging, 17*(3), 443–452.

Aspden, P., Wolcott, J. A., Bootman, J. L., & Croenwett, L. R. (2007). *Preventing medication errors*. Washington, DC: The National Academies Press.

Azevedo, R. F. L., Morrow, D. G., Hasegawa-Johnson, M., Gu, K., Soberal, D., Huang, T., et al. (2015). Improving patient comprehension of numeric health information. In *Proceedings of the human factors and ergonomics society 59th annual meeting*. Santa Monica: Human Factors & Ergonomics Society.

Baek, J., & Seidman, R. L. (2015). Impact of information technology, clinical resource constraints, and patient-centered practice characteristics on quality of care. *Health Services Research and Managerial Epidemiology, 2*. http://dx.doi.org/10.1177/2333392815572340.

Baltes, P. B., & Baltes, M. M. (1990). Psychological perspectives on successful aging: the model of selective optimization with compensation. *Successful Aging: Perspectives from the Behavioral Sciences, 1*, 1–34.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening, 25*, 178–198.

Beier, M. E., & Ackerman, P. L. (2005). Age, ability and the role of prior knowledge on the acquisition of new domain knowledge. *Psychology and Aging, 20*, 341–355.

Bickmore, T., Pfeifer, L., Byron, D., Forsythe, S., Henault, L., Jack, B., et al. (2010). Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. *Journal of Health Communication, 15*, 197–210.

Bloom, R. L., Obler, L. K., De Santi, S., & Ehrlich, J. S. (2013). *Discourse analysis and applications: Studies in adult clinical populations*. Psychology Press.

Budnitz, D. S., & Layde, P. M. (2007). Outpatient drug safety: new steps in an old direction. *Pharmacoepidemiology and Drug Safety, 16*, 160–165.

Cardosi, K., Falzarano, P., & Han, S. (1998). *Pilot-controller communication errors: An analysis of aviation safety reporting system (ASRS)*. Reports (No. DOT-VNTSC-FAA-98-4). Cambridge MA: Volpe National Transportation Systems Center.

Carr, N. (2011). *The Shallows: What the Internet is doing to our brains*. WW Norton & Company.

Carroll, J. S., Williams, M., & Gallivan, T. M. (2012). The ins and outs of change of shift handoffs between nurses: a communication challenge. *BMJ Quality and Safety, 21*, 586–593.

Chin, J., Madison, A., Stine-Morrow, E. A. L., Gao, X., Graumlich, J. F., Murray, M. D., et al. (2015). Cognition and health literacy in older adults' recall of self-care information. *The Gerontologist*, 1–9. http://dx.doi.org/10.1093/geront/gnv091. Advance Access publication July 24, 2015.

Chin, J., Morrow, D., Stine-Morrow, E. A. L., Conner-Garcia, T., Graumlich, J. F., & Murray, M. D. (2011). The process-knowledge model of health literacy: evidence from a componential analysis of two commonly used measures. *Journal of Health Communication, 16*, 222–241.

Chin, J., Payne, B., Gao, X., Conner-Garcia, T., Graumlich, J., Murray, M. D., et al. (2015). Memory and comprehension for health information among older adults: distinguishing the effects of domain-general and domain-specific knowledge. *Memory, 23*, 577–589.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington DC: American Psychological Association.

Coiera, E. (2006). Communication systems in healthcare. *Clinical Biochemistry Review, 27*, 89–98.

Collins, S. A., Bakken, S., Vawdrey, D. K., Coiera, E., & Currie, L. (2011). Clinician preferences for verbal communication compared to EHR documentation in the ICU. *Applied Clinical Informatics, 2*, 190.

Conner-Garcia, T., Morrow, D. G., Graumlich, J. F., Ellison, J., & Wang, H. (2012). Outpatient perceptions of the Medtable: a medication scheduling tool. *The Journal of Pharmacy Technology, 28*, 163–170.

Cordero, C. (2011). *Advancing effective communication, cultural competence, and patient-and family-centered care: A roadmap for hospitals*. The Joint Commission and National Association of Public Hospitals and Health Systems.

Craik, F. I. (1994). Memory changes in normal aging. *Current Directions in Psychological Science, 3*, 155–158.

Daselaar, S. M., Fleck, M. S., Dobbins, I. G., Madden, D. J., & Cabeza, R. (2006). Effects of healthy aging on hippocampal and rhinal memory functions: an event-related fMRI study. *Cerebral Cortex, 16*, 1771–1782.

Day, R. S. (1998). Alternative representations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 261–305). New York: Academic Press.

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied, 14*, 392–400.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory and Cognition, 25*, 691–700.

Ekeocha, J. O., & Brennan, S. E. (2008). Collaborative recall in face-to-face and electronic groups. *Memory, 16*, 245–261.

Ertl, B., Kopp, B., & Mandl, H. (2008). Supporting learning using external representations. *Computers and Education, 51*, 1599–1608.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*, 213–232.

Friedman, R. H., Stollerman, J., Rozenblyum, L., Belfer, D., Selim, A., Mahoney, D., et al. (1998). A telecommunications system to manage patients with chronic disease. *Studies in Health Technology Information, 52*, 1330–1334.

Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *American Journal of Public Health, 99*, 2196–2202.

Gawande, A. (2010). *The checklist manifesto: How to get things right*. New York: Metropolitan Books.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Action as language in a shared visual space. In *Proceedings of computer supported cooperative work* (pp. 487–496). New York: ACM Press.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: pictures help to build mental models. *Journal of Memory and Language, 31*, 129–151.

Gray, W. D., & Fu, W. T. (2004). Soft constraints in interactive behavior: the case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*, 359–382.

Haidet, P. (2007). Jazz and the 'art' of medicine: improvisation in the medical encounter. *The Annals of Family Medicine, 5*, 164–169.

Hazlehurst, B., Gorman, P. N., & McMullen, C. K. (2008). Distributed cognition: an alternative model of cognition for medical informatics. *International Journal of Medical Informatics, 77*, 226–234.

Henkel, L. A., & Rajaram, S. (2011). Collaborative remembering in older adults: age-invariant outcomes in the context of episodic recall deficits. *Psychology and Aging, 26*, 532–545.

Horton, W., & Spieler, D. (2007). Age-related differences in communication and audience design. *Psychology and Aging, 22*, 281.

Huang, T. S., Hasegawa-Johnson, M. A., Chu, S. M., Zeng, Z., & Tang, H. (2009). Sensitive talking heads [Applications corner]. *Signal Processing Magazine, IEEE, 26*, 67–72.

Hupet, M., Chantraine, Y., & Nef, F. (1993). References in conversation between young and old normal adults. *Psychology and Aging, 8*, 339–346.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19*, 265–288.

Ingersoll, K.. S., & Cohen, J. (2008). The impact of medication regimen factors on adherence to chronic treatment: a review of literature. *Journal of Behavioral Medicine, 31*, 213–224.

Insel, K. C., Morrow, D. G., Brewer, B. B., & Figuerosa, A. J. (2006). Cognitive function and medication adherence. *Journal of Gerontology: Psychological Sciences, 61B*, P102—P107.

Institute of Medicine. (2012). *Health IT and patient safety: Building safer systems for better care*. Washington DC: The National Academies Press.

Joint Commission on Accreditation of Healthcare Organizations. (2005). *Root causes of medication errors (1995—2004)*. Retrieved 11.02.05.

Kannampallil, T., Waicekauskas, K., Morrow, D., Kopren, K., & Fu, W.-T. (2013). Collaborative tools for a simulated patient-provider medication scheduling task. *Cognition, Technology, and Work, 15*, 121—131.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Larkin, J. H., & Simon, H.. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65—100.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: The MIT Press.

Liu, L. L., & Park, D. C. (2004). Aging and medical adherence: the use of automatic processes to achieve effortful things. *Psychology and Aging, 19*, 318—325.

Mayer, R. E., & DaPra, C. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied, 3*, 239—252.

Meade, M. L., Nokes, T. J., & Morrow, D. G. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory, 17*, 39—48.

Monk, A. F. (2009). Common ground in electronically mediated conversation. In J. M. Carroll (Ed.), *Synthesis lectures on human-centered informatics #1* (pp. 1—45). Morgan & Claypool.

Morrow, D. G., Carver, L. M., Leirer, V. O., & Tanke, E. D. (2000). Medication schemas and memory for automated telephone messages. *Human Factors, 42*, 523—540.

Morrow, D. G., & Chin, J. (2012). Technological bridges between health care systems and older adults. In R. Hill, R. Zheng, & M. Gardner (Eds.), *Engaging older adults with modern technology: Internet use and information access needs* (pp. 59—79). Hershey, PA: IGI Publishing.

Morrow, D. G., & Chin, J. (2015). Health literacy and health decision making among older adults. In T. Hess, J. Strough, & C. Lockenhoff (Eds.), *Aging and decision-making: Empirical and applied perspectives* (pp. 261—282). London: Elsevier.

Morrow, D. G., Conner-Garcia, T., Graumlich, J., Wolf, M., McKeever, S., Madison, A., et al. (2012). An EMR-based tool to support collaborative planning for medication use among adults with diabetes: a multi-site randomized control Trial. *Contemporary Clinical Trials, 33*, 1023—1032.

Morrow, D. G., & Czaja, S. J. (2015). The implications of aging for human-system integration. In D. Boehm-Davis, F. Durso, & J. D. Lee (Eds.), *Handbook of human-system integration* (pp. 535—552). Washington, DC: American Psychological Association.

Morrow, D. G., & Durso, F. T. (2011). Patient safety research that works: introduction to the special issue on human performance and health care. *Journal of Experimental Psychology: Applied, 17*, 191—194.

Morrow, D. G., & Fischer, U. M. (2013). Communication in socio-technical systems. In J. D. Lee, & A. Kirlik (Eds.), *Oxford handbook of cognitive engineering* (pp. 178—199). New York, NY: Oxford University Press.

Morrow, D. G., Hier, C., Menard, W. E., & Leirer, V. O. (1998). Icons improve older and younger adult comprehension of medication information. *Journal of Gerontology B: Psychological Science, 53B*, 240—254.

Morrow, D. G., & Leirer, V. O. (2001). A patient-centered approach to designing automated telephone messages for older adults. In W. Rogers, & D. Fisk (Eds.), *Human factors interventions for the health care of older adults* (pp. 179—203). Mahwah, NJ: Erlbaum.

Morrow, D. G., Leirer, V. O., Carver, L. M., & Tanke, E. D. (1998). Older and younger adult memory for health appointment information: implications for automated telephone messaging design. *Journal of Experimental Psychology: Applied, 4*, 352—374.

Morrow, D. G., Leirer, V. O., Carver, L. M., Tanke, E. D., & McNally, A. (1999a). Repetition improves older and younger adult memory for automated appointment messages. *Human Factors, 41*, 194—204.

Morrow, D. G., Leirer, V. O., Carver, L. M., Tanke, E. D., & McNally, A. (1999b). Effects of aging, message repetition, and note-taking on memory for health information. *Journal of Gerontology: Psychological Sciences, 54B*, P369—P379.

Morrow, D. G., Raquel, L. M., Schriver, A. T., Redenbo, S., Rozovski, D., & Weiss, G. (2008). External support for collaborative problem solving in a simulated provider/patient medication scheduling task. *Journal of Experimental Psychology: Applied, 14*, 228—297.

Morrow, D. G., Ridolfo, H. E., Menard, W. E., Sanborn, A., Stine-Morrow, E. A. L., Magnor, C., et al. (2003). Environmental support promotes expertise-based mitigation of age differences in pilot communication tasks. *Psychology and Aging, 18*, 268—284.

Morrow, D., Rodvold, M., & Lee, A. (1994). Nonroutine transactions in controller-pilot communication. *Discourse Processes, 17*, 235—258.

Morrow, D. G., & Rogers, W. A. (2008). Environmental support: an integrative framework. *Human Factors, 50*, 589—613.

Morrow, D. G., Wickens, C. D., Rantanen, E. M., Chang, D., & Marcus, J. (2008). Designing external aids that support older pilots' communication. *International Journal of Aviation Psychology, 18*, 167—182.

Morrow, D. G., & Wilson, E. A. H. (2010). Medication adherence among older adults: a systems perspective. In J. C. Cavanaugh, & C. K. Cavanaugh (Eds.), *Aging in America: Psychological, physical, and social issues* (Vol. 2). Westport CT: Greenwood.

National Research Council. (2012a). Education for life and work: developing transferable knowledge and skills in the 21st century. In J. W. Pellegrino, & M. L. Hilton (Eds.), *Committee on defining deeper learning and 21st century skills*. Washington, DC: The National Academies Press.

National Research Council. (2012b). Improving adult literacy instruction: options for practice and research. In A. M. Lesgold, & M. Welch-Ross (Eds.), *Committee on learning sciences: Foundations and applications to adolescent and adult literacy*. Washington, DC: The National Academies Press.

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics, 55*, 1059—1072.

Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., & Stine-Morrow, E. A. L. (2012). The effects of print exposure on sentence processing and memory in older adults: evidence for efficiency and reserve. *Aging, Neuropsychology, and Cognition, 19*, 122—149.

Peters, E., Dieckmann, N. F., Västfjäll, D., Mertz, C. K., Slovic, P., & Hibbard, J. H. (2009). Bringing meaning to numbers: the impact of evaluative categories on decisions. *Journal of Experimental Psychology: Applied, 15*(3), 213—227.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation, 4*, 203—228.

Piette, J. D., Rosland, A., Marinec, N. S., Striplin, D., Bernstein, S. J., & Silveira, M. J. (2013). Engagement with automated patient monitoring and self-management support calls: experience with a thousand chronically-ill patients. *Medical Care, 51*, 216—223.

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin and Review, 14*, 1027—1042.

Reyna, V. (2011). Across the life span. In B. Fischoff, N. Brewer, & J. Downs (Eds.), *Communicating risks and benefits: A users guide* (pp. 111—120). Washington DC: FDA.

Rubin, R. D., Brown-Schmidt, S., Duff, M. C., Tranel, D., & Cohen, N. J. (2011). How do I remember that I know you know that I know? *Psychological Science, 22*(12), 1574–1582, 0956797611418245.

Schillinger, D., Piette, J., Grumbach, K., Wang, F., Wilson, C., Daher, C., et al. (2003). Closing the loop: physician communication with diabetic patients who have low health literacy. *Archives of Internal Medicine, 163*, 83–90.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences, 4*, 321–354.

Schwartzberg, J. G., Cowett, A., VanGeest, J., & Wolf, M. S. (2007). Communication techniques for patients with low health literacy: a survey of physicians, nurses, and pharmacists. *American Journal of Health Behavior, 31*, S96–S104.

Shannon, C. E., & Weaver, N. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Siek, K. A., Khan, D. U., Ross, S. E., Haverhals, L. M., Meyers, J., & Cali, S. R. (2011). Designing a personal health application for older adults to manage medications: a comprehensive case study. *Journal of Medical Systems, 35*, 1099–1121.

Stead, W. W., & Linn, H. S. (2009). *Computational technology for effective health care: Immediate steps and strategic directions*. Washington, DC: National Academies Press.

Stewart, M. (1995). Effective physician-patient communication and health outcomes: a review. *Journal of the Canadian Medical Association, 152*, 1423–1433.

Stine-Morrow, E. A., & Miller, L. (2009). Aging, self-regulation, and learning from text. *Psychology of learning and motivation, 51*, 255–296.

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.

Taha, J., Czaja, S., Sharit, J., & Morrow, D. (2013). Factors affecting the usage of a personal health record (PHR) to manage health. *Psychology and Aging, 28*, 1124–1139.

Tang, P. C., Ash, J. S., Bates, D. W., Overhage, M., & Sands, D. Z. (2006). Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association, 13*, 121–126.

Tarn, D. M., Heritage, J., Paterniti, D. A., Hays, R. D., Kravitz, R. L., & Wenger, N. S. (2006). Physician communication when prescribing new medications. *Archives of Internal Medicine, 166*, 1855–1862.

Tarn, D. M., Paterniti, D. A., Kravitz, R. L., Fein, S., & Wenger, N. S. (2009). How do physicians conduct medication reviews? *Journal of General Internal Medicine, 24*, 1296–1302.

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., Hendriks, M., & Schmidt, H. G. (2003). The efficiency of multimedia learning into old age. *British Journal of Educational Psychology, 73*, 489–505.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1160–1175.

White, A., & Danis, M. (2013). Enhancing patient-centered communication and collaboration by using the electronic health record in the examination room. *JAMA, 309*, 2327–2328.

Whitten, P. (2001). The state of telecommunication technology to enhance older adults' access to health services. In W. A. Rogers, & A. D. Fisk (Eds.), *Human factors interventions for the health care of older adults* (pp. 121–146). Mahwah, NJ: Erlbaum.

Wickens, C. D., & Hollands, J. (1992). *Engineering psychology and human performance*. Glenview , IL: Scott, Foreman, and Co.

Wolf, M.. S., Curtis, L. M., Waite, K., Bailey, S. C., Hedlund, L. A., Davis, T. C., et al. (2011). Helping patients simplify and safely use complex medication regimens. *Archives of Internal Medicine, 171*, 300–305.

Young, A., Chaudhry, H. J., Rhyne, J., & Dugan, M. (2011). A census of actively licensed physicians in the United States, 2010. *Journal of Medical Regulation, 96*, 10–20.
Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*, 1–11.

This page intentionally left blank

INDEX

Note: Page numbers followed by "f" indicate figures and "t" indicate tables.

E

F

G

H

I

K

This page intentionally left blank

CONTENTS OF PREVIOUS VOLUMES

VOLUME 40

VOLUME 41

VOLUME 42

VOLUME 43

VOLUME 44

VOLUME 45

VOLUME 46

VOLUME 47

VOLUME 48

VOLUME 49

VOLUME 50

VOLUME 51

VOLUME 52

VOLUME 53

VOLUME 54

VOLUME 55

VOLUME 56

VOLUME 57

VOLUME 58

VOLUME 59

VOLUME 60

VOLUME 61

VOLUME 62

VOLUME 63

This page intentionally left blank

ELSEVIER

ACADEMIC PRESS

An imprint of Elsevier
store.elsevier.com

ISBN 978-0-12-804739-2

9 780128 047392