Pillars of judgment: How memory abilities, task feedback, and cognitive load guide
judgment strategies

Inauguraldissertation
zur
Erlangung der Würde
eines Doktors der Philosophie
vorgelegt der
Fakultät für Psychologie
der Universität Basel

von

Janina Anna Hoffmann

aus Böblingen, Deutschland

Basel, 2014

UNI
BASEL

Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. Jörg Rieskamp
Prof. Dr. Stephan Lewandowsky

Basel, den _____

<div align="right">

_____
Prof. Dr. Roselind Lieb

</div>

UNI
BASEL

Declaration

I, Janina A. Hoffmann (born June 16th, 1986 in Böblingen, Germany) hereby declare the following:

(i)   My cumulative dissertation is based on three manuscripts, **one published, one under revision, and one submitted**. I contributed substantially and independently to all manuscripts in this dissertation and have been primarily responsible for the ideas, data collection, analyses, and writing of the papers. This characterization of my contributions is in agreement with my co-authors' views.

(ii)   I only used the resources indicated.

(iii)   I marked all the citations.

Basel, 30th May 2014

Janina A. Hoffmann

Acknowledgements

Finding the right strategy to start, pursue, and finish writing a dissertation is not always an easy task. First and foremost, I would like to thank those two persons who always kept me on the right track: my advisors, Bettina von Helversen and Jörg Rieskamp, for their tremendous support, advice, and help. It has been a pleasure to work with you on inspiring scientific questions and I appreciate what I have learned from you over the past few years.

I would also like to thank my colleagues at the Department of Economic Psychology, especially Sandra Andraszewicz who helped me to slowly transform our office into a living room. I am indebted to my former colleagues from the University of Mannheim and Plymouth. I am also grateful to our research assistants, especially Mélanie Moos and Regina Weilbächer, for their precious help as well as to Laura Wiles and Anita Todd for editing the manuscripts. Finally, I thank Sandra Andraszewicz and Nicolas Berkowitsch for reviewing a draft of my framework.

I would like to thank my family for their love and their support, especially my twin sister Natalie for lighting up my life with elephants, donkeys, and frog kings making it "Zwei-kleine-Wölfe-schön".

Finally, I would like to thank those people who made even rainy days sunnier, brightened up grey skies, and colored my life in the last three years: Camille Cotillon, Julia Sowislo, Sasha Lemola, and Rosanne Kirchberg for unforgettable coffee-beer-wine-chatting-dancing sessions, Katrin Heil, Manuela Michelbach, Marius Hoffmann, Jana Porten, and Stephi Nähr for reminding me of the child I still am, Selma Rudert and Sandra Marg for the most compelling ideas to rule the world, Sonja Wagner, Nico Herrmann, Vera Weinbeer, Gerald Oeckl, Stefan Rapp, Daniela Borg, and Natascha Jansen for the greatest dinners and most relaxing city trips, Thomas Rösler for always bringing me safely home, and Stephan Vogel for fixing all nails I screwed up.

Abstract

Making judgments is an essential part of everyday life and how people form a judgment has instigated a plethora of research. Research in judgment and categorization has particularly contrasted two types of judgment strategies: rule-based and similarity-based strategies. Recent research suggests that people can make use of both rule- and similarity-based strategies and frequently shift between these strategies. To select between strategies, contingency approaches propose that people trade off the strategies' accuracy against the effort needed to execute strategy so that the selected strategy matches the demands of the task environment and the capabilities of the decision maker. This dissertation presents three papers investigating how accuracy-effort trade-offs between rule-based and similarity-based judgment strategies change strategy selection in judgment and categorization tasks.

The first paper studies how reducing working memory by imposing a cognitive load may foster shifts to a less demanding similarity-based strategy and, in turn, enhances judgment performance in tasks well solved by a similarity-based strategy, but not in tasks for which rules are better suited. The second paper compares judgment strategies to strategies people apply in categorization. It shows that the same task characteristics, namely the number of cues and the functional relationship between cues and criterion, foster shifts between rule-based and similarity-based strategies in judgment and categorization. The third manuscript explores which memory abilities underlie rule-based and similarity-based judgments. Specifically, it shows that working memory predicts to a stronger degree how well people solve rule-based judgment tasks, whereas episodic memory is more closely linked to judgment performance in similarity-based tasks. Furthermore, episodic memory also predicts selecting a similarity-based strategy, but not working memory.

Strolling through a typical bookstore, one quickly notices that shelves are covered with books called *The Art of Thinking Clearly*, *Allen Carr's Easy Way to Stop Smoking*, *What Women Really Want in Bed,* or *How to Cook Everything.* In 2012, the sale of guidebooks in fact made up 13.8% of the total book sales in Germany (Börsenverein des Deutschen Buchhandels, 2013). Offering a way of solving daily life problems apparently meets the demands of the readers. In daily life, however, there are often variable routes to success: Spontaneously asking someone for his phone number may succeed in a bar, but seems to be an inappropriate pick-up strategy in an art gallery. In contrast, starting a philosophical discussion may pique someone's interest in a gallery, but in a noisy club the discussion will probably be overheard. Hence, whether the strategy one follows is crowned with success often depends on the context or task environment (Beilock & DeCaro, 2007; Markman, Maddox, & Worthy, 2006).

In recent decades, the idea that people possess a repertoire of strategies flourished in different fields of psychology ranging from memory (Dunlosky & Kane, 2007; McNamara & Scott, 2001) to categorization (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & O'Brien, 2005; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994), judgment (Juslin, Karlsson, & Olsson, 2008; Juslin, Olsson, & Olsson, 2003; von Helversen & Rieskamp, 2008, 2009), and decision making (Beach & Mitchell, 1978; Gigerenzer, Todd, & the ABC Research Group, 1999; Payne, Bettman, & Johnson, 1988, 1993). The concept of the adaptive toolbox, for instance, assumes that individuals can be characterized by a set of cognitive mechanisms that exploit evolved capacities (Gigerenzer et al., 1999; Gigerenzer, Hoffrage, & Goldstein, 2008; Goldstein & Gigerenzer, 2011). Similarly, theories in categorization have repeatedly argued that people can rely upon qualitatively different categorization strategies that build upon distinct memory systems (Ashby & O'Brien, 2005; Sloman, 1996; Smith & Grossman, 2008; Smith, Patalano, & Jonides, 1998). However, the idea that individuals may apply several strategies to solve problems opens up the question of how people select among these different strategies (Marewski & Schooler, 2011; Rieskamp &

Otto, 2006). One early solution to this strategy selection problem has been offered by contingency approaches to judgment and decision making (Beach & Mitchell, 1978; Payne et al., 1993).

Contingency approaches portray the decision maker as actively selecting strategies that are adapted to the task environment and the cognitive capabilities of the decision maker (Beach & Mitchell, 1978; Payne et al., 1993). According to this approach, strategy selection constitutes a compromise between the accuracy achieved by using a strategy and the effort of executing a strategy. Selecting a particular strategy presupposes that the strategy is available in a person's strategy repertoire (Beach & Mitchell, 1978; Lemaire & Siegler, 1995). Second, the person needs to know that this strategy is applicable to the decision problem at hand (Beach & Mitchell, 1978; Lemaire & Siegler, 1995). The task environment offers feedback about the strategies' accuracy and thereby increases the likelihood of selecting appropriate strategies and diminishes the likelihood of following inappropriate strategies (Rieskamp & Otto, 2006). Third, the decision maker needs to be willing and able to execute the strategy correctly (Beach & Mitchell, 1978; Lemaire & Siegler, 1995). Time pressure, for instance, limits the time available for executing a strategy and hence may force the individual to apply simplifying strategies (Wright, 1974). Likewise, distractions impose an additional cognitive load on the decision maker and may restrict how much effort the decision maker can invest in strategy execution (Beach & Mitchell, 1978). Finally, learning about the strategies' benefits and costs may strengthen — over time — individual preferences for applying specific strategies and these stable tendencies may be linked to cognitive abilities (Bröder, 2003) or age (Mata, von Helversen, Karlsson, & Cüpper, 2012).

This dissertation contributed to the problem of strategy selection by investigating how task demands and memory abilities affect strategy use in judgment problems and how these judgment strategies, in turn, facilitate or impede judgments depending on the task environment. In the first manuscript, we focused on how reducing working memory capacity

fosters shifting to less demanding strategies and how this shift may help performance in judgment tasks for which the less demanding strategy is better suited. In the second manuscript we studied whether the same task components affect strategy selection in judgment and categorization by systematically varying the number of cues and the functional relationship between cues and criterion. The third manuscript, finally, focused on the question of how memory abilities promote judgments by facilitating strategy choice and strategy execution. In all studies, we put our emphasis on contrasting two kinds of judgment strategies: rule-based and similarity-based judgment strategies.

Judgment Strategies

People encounter judgment problems every day ranging from considering the suitability of a business dress to judging the attractiveness of an apartment to evaluating the effectiveness of a political program. Coming up with such a judgment requires inferring a continuous criterion, for instance the apartment's attractiveness, from a number of critical attributes of this object (the cues), such as the size of the apartment or the monthly rent.

Cognitive science particularly has contrasted two kinds of strategies (or cognitive processes[1]): rule-based and similarity-based strategies (Erickson & Kruschke, 1998; Juslin et al., 2003; Nosofsky et al., 1994; von Helversen & Rieskamp, 2008, 2009). On the one hand, the decision maker may abstract rules describing how each cue relates to the criterion and find out the importance of each cue. To make a judgment, rule-based strategies assume that people finally combine the weighted cue values in an additive fashion (Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Juslin et al., 2003). For instance, a tenant looking for a new apartment may try to figure out how much he appreciates a large apartment or a modern kitchen and assign a high weight to the apartment's size. Consequently, the tenant will rate large apartments more favorably. Linear additive models have been predominantly used to capture these rules

---

[1] The terms "processes" and "strategies" will be used interchangeably in this dissertation.

(Cooksey, 1996) and can describe people's judgments in a variety of tasks (Brehmer & Brehmer, 1988) ranging from personal selection (Graves & Karren, 1992) to medical diagnoses (Wigton, 1996).

On the other hand, the tenant may follow a similarity-based strategy and judge the apartment's attractiveness by comparing it to past apartments he lived in. Similarity-based strategies, such as the exemplar model (Juslin et al., 2003, 2008; Medin & Schaffer, 1978; Nosofsky, 1988), have been successfully applied to various areas in psychology from categorization to reasoning to memory. Exemplar models rely upon the retrieval of past experiences from long-term memory assuming that all previously encountered objects (exemplars) are stored in memory along with their criterion values (Juslin et al., 2003, 2008). To judge the new object (the probe), previously encountered exemplars are retrieved from memory. The more similar a retrieved exemplar is to the probe, the more it influences the final judgment. Accordingly, if a tenant has already lived in an apartment with a similar floor plan, he may just recall how much he enjoyed living in his former apartment to rate the suitability of the new apartment.

Obviously, the distinction between similarity and rules is at the heart of cognitive science (Hahn & Chater, 1998; Pothos, 2005; Sloman, 1996). The reason why various fields have repeatedly contrasted these strategies is that similarity- and rule-based strategies offer two fundamentally distinct ways of representing knowledge (Hahn & Chater, 1998; Juslin et al., 2003): Whereas similarity-based processes base inferences upon concrete instances stored in memory, rule-based processes rely upon explicit knowledge abstraction. Past research suggests that humans can rely upon both kinds of processes (Erickson & Kruschke, 1998; Juslin et al., 2003; von Helversen & Rieskamp, 2008, 2009); indeed, within a specific domain such as judgment, the conditions triggering rule-based or similarity-based strategies are better and better understood (Karlsson, Juslin, & Olsson, 2008).

Accuracy-Effort Trade-offs in the Selection of Judgment Strategies

Accumulating evidence suggests that people adapt the judgment strategy to the task at hand (Juslin et al., 2003, 2008; Karlsson, Juslin, & Olsson, 2007; Karlsson et al., 2008). One major factor shaping strategy selection is the relative accuracy that can be reached by executing rule-based or similarity-based strategies. Rule-based models can capture judgments well in linear multiple-cue judgment tasks in which the criterion is a linear additive function of the cues (Juslin et al., 2003, 2008). In multiplicative judgment tasks, however, the criterion is a multiplicative function of the cues and, thus, task feedback strongly discourages rule abstraction processes, because a linear additive model cannot well represent the relationship between the cues and the criterion (Juslin et al., 2008). Therefore, people should shift to exemplar memory. Confirming this idea, it has been consistently found that more people rely on similarity-based processes in multiplicative judgment tasks (Juslin et al., 2008; Karlsson et al., 2007). Not all accuracy-based strategy shifts are necessarily successful on the first attempt. In nonlinear judgment tasks, for instance, a similarity-based strategy may not lead instantaneously to a good performance and so people shift back to the default, but inappropriate rule-based strategy (Karlsson et al., 2008; Olsson, Enkvist, & Juslin, 2006). Accordingly, Karlsson et al. (2008) argued that executing a similarity-based strategy requires a deliberative strategic choice.

Another major factor shaping strategy selection is the effort associated with executing rule-based and similarity-based strategies. Time pressure, for instance, has been found to reduce the consistency with which individuals implement a linear judgment policy in nonlinear judgment tasks (Rothstein, 1986). In a similar vein, cognitive load impairs rule-based strategies more than implicit or similarity-based strategies suggesting that people may shift more to similarity-based strategies under cognitive load (Filoteo, Lauritzen, & Maddox, 2010; Juslin et al., 2008; Zeithamova & Maddox, 2006). In contrast, if abstraction of linear rules is facilitated, for instance, by only changing one cue between trials or because the cue

directions are known, more people rely on rule-based learning (Juslin et al., 2008; Platzer & Bröder, 2013; von Helversen, Karlsson, Mata, & Wilke, 2013). Finally, feedback can also render rule abstraction more difficult and, hence, increase the effort of following a rule-based strategy. For instance, binary feedback in categorization often leads to switches to a similarity-based strategy because diminished feedback quality makes abstracting the correct rule more difficult (Juslin et al., 2003; Karlsson et al., 2008; von Helversen et al., 2013).

How accurate and effortful certain strategies are may be learned over time (Rieskamp & Otto, 2006). Consequently, people may build up stable tendencies for rule-based or similarity-based learning that may be related to stable personal characteristics such as memory abilities (McDaniel, Cahill, Robbins, & Wiener, 2013). For instance, people with good episodic memory may prefer applying a similarity-based strategy. In this vein, it has been found that older adults are less likely to follow a similarity-based strategy— possibly, because they do not trust their long-term memory (Mata, von Helversen, et al., 2012). Alternatively, also adaptively choosing a strategy may hinge upon memory abilities (Mata, Pachur et al., 2012). In this spirit, higher intelligence helps to ignore information in case ignorance is adaptive (Bröder, 2003). Likewise, high working memory capacity does not predict which strategy people choose, but how good they are at following it (Craig & Lewandowsky, 2012; Lewandowsky, Yang, Newell, & Kalish, 2012).

Manuscript 1 particularly investigated how increasing the difficulty of abstracting rules by introducing a cognitive load can foster similarity-based judgment strategies and — depending on the accuracy that can be achieved by relying upon exemplar memory— can even benefit performance. Manuscript 2 pronounces how effort, manipulated by the number of cues, and accuracy, manipulated by the functional relationship between cues and criterion, interact to reinforce rule-based and similarity-based strategies across judgment and categorization tasks. Finally, Manuscript 3 goes one step further by investigating how individual differences in strategy use and judgment accuracy are grounded in memory

abilities thus shifting the focus even more towards how stable personal characteristics may be

linked to preferences in strategy use.

Making Judgments Under Cognitive Load

Distractions, such as a phone call from a student while writing your dissertation, are a hassle

in daily life and often disturb performance. Distractions hurt performance because they

impose an additional working memory load on the decision maker (Baddeley & Hitch, 1974).

Accordingly, under cognitive load, people tend to shift to strategies that are less working

memory demanding, but often also less accurate (Juslin, et al., 2008; Payne, et al., 1993). In

our paper, we suggested that under some circumstances this shift could also be beneficial for

performance — in cases when the less demanding strategy provides a better solution to the

problem at hand (Beilock & DeCaro, 2007).

To test this hypothesis, our participants learned to solve a judgment task under a high

cognitive load, a low cognitive load, or without cognitive load. This judgment task could be

solved better by either a similarity-based judgment strategy (Experiment 1) or a rule-based

judgment strategy (Experiment 2). While rule-based strategies should draw highly upon

working memory capacity and rule abstraction is severely impaired under cognitive load

(Filoteo et al., 2010; Juslin et al., 2008; Zeithamova & Maddox, 2006), similarity-based

strategies may rely to a lesser extent upon working memory capacity and may be rather driven

by implicit, associative processes (Sloman, 1996). Accordingly, under cognitive load, people

should abandon a rule-based strategy more often and shift to the less demanding similarity-

based strategy. In Experiment 1, we tested, whether this shift proves beneficial for judgment

performance in a multiplicative judgment task that can be better be solved by a similarity-based strategy.

In a training phase, our participants first learned to judge on a continuous scale how many small creatures different comic figures could catch. To predict the criterion, people could use five different features (or cues) of the comic figure (e.g., the shape of the ears). While judging these comic figures, participants had to remember two, four, or no letters to induce cognitive load. After each trial, participants received feedback about their judgment accuracy. This training phase finished when participants reached a learning criterion or the maximum number of training blocks. Afterwards, participants moved to a test phase in which they judged known and unknown comic figures twice without getting any performance feedback and without a concurrent cognitive load.

At the end of the training phase, judgment accuracy did not differ between participants learning under high, low or without cognitive load. However, in the test phase, increasing cognitive load helped participants to make more accurate judgments for unknown items. To analyze more closely why performance even improved under cognitive load, we fitted three different cognitive models — a linear rule-based model, a similarity-based exemplar model, and a baseline model — to participants' judgments at the end of training and predicted participants' judgment in the test phase with the fitted parameters (a generalization test; Busemeyer & Wang, 2000). Under high cognitive load, more participants were better described by an exemplar model than by a linear or a baseline model. Moreover, this shift to a similarity-based strategy mediated the effect of cognitive load on judgment performance.

Cognitive load, hence, increased shifting to similarity-based strategies and, in turn, improved judgment accuracy. However, shifting to a similarity-based strategy may harm performance in a judgment task that can best be solved by using rules. In Experiment 2, we tested how cognitive load affects strategy use and performance in a linear judgment task that can best be solved by more demanding rule abstraction strategy. Replicating a study from

Mata, von Helversen et al. (2012), participants learned to solve a linear judgment under a high cognitive load or without cognitive load. As in Experiment 1, under load participants switched more to similarity-based strategies, but this shift was less pronounced. In the linear task, however, following a similarity-based strategy harmed judgment accuracy for unknown items.

In sum, increasing cognitive load makes rule abstraction more difficult and increases reliance upon less demanding similarity-based strategies. In addition, increasing cognitive load does not lead per se to worse performance, but can sometimes even improve performance — depending upon how well the less demanding strategy matches the problem at hand.

Strategy Shifts in Judgment and Categorization

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013b). From rules to exemplars: Similar task features shape judgment and categorization processes. Manuscript to be submitted for publication.

In some college courses, teachers are asked to judge students' essays on a continuous grading scale — a typical judgment task; in other courses, however, teachers are only asked to categorize their students into the categories "pass" or "fail" — a usual categorization task. How teachers grade their students should obviously not depend on the response scale: the literature, however, has seldom linked judgment strategies to categorization strategies and vice versa (Juslin et al., 2003). On the one hand, rule-based and similarity-based strategies have been indeed proposed to underlie both judgments and categorizations (Erickson & Kruschke, 1998; Juslin et al., 2003, 2008; Nosofsky et al., 1994). On the other hand, people frequently shift from rule-based judgment strategies to similarity-based categorization strategies (Juslin et al., 2003; Pachur & Olsson, 2012; von Helversen, Mata, & Olsson, 2012;

von Helversen et al., 2013) and task characteristics identified as fostering shifts from rule-based to similarity-based strategies vary between judgment and categorization. Whereas categorization research has intensively studied how the number of cues affects strategy choice (e.g. Ashby, Maddox, & Bohil, 2002; Filoteo et al., 2010; Maddox & Ashby, 2004; Zeithamova & Maddox, 2006), judgment research has pronounced the importance of the functional relationship in strategy selection (Hoffmann et al., 2013a; Juslin et al., 2008; Karlsson et al., 2007). Consequently, it is still unclear whether strategy shifts from judgment to categorization generalize across a variety of task characteristics.

To integrate the fields of judgment and categorization, we investigated how the number of cues and the functional relationship between cues and criterion affect strategy choice across the same categorization and judgment task. Overall, more cues (or dimensions) may increase the effort associated with rule abstraction (Karelaia & Hogarth, 2008), whereas complex functional relationships, such as multiplicative functions, cannot be learned by abstracting linear rules (Juslin et al., 2008). Accordingly, a higher number of cues and more complex functional relationships should increase reliance upon exemplar memory in judgment and categorization. Furthermore, if the strategies people use to categorize objects match their judgment strategies people should rely upon similar strategies in both tasks (McDaniel et al., 2013). We investigated this question in two experiments in which participants solved both a categorization and a judgment task with the same underlying task structure. In two experiments we varied the task structure from a one-dimensional linear rule predicting judgments and category membership to a multidimensional linear rule to a multidimensional, multiplicative function (Experiment 1) and extended this to a multidimensional quadratic function (Experiment 2). In a training phase, participants learned to predict the judgment criterion or the category, respectively, of 25 objects based on four continuous features. After each trial, participants received feedback on their performance. In the subsequent test phase, participants judged or categorized 15 new objects four times. We

analyzed judgment and categorization strategies by using a generalization test to classify participants as following a rule-based strategy, a similarity-based strategy or a baseline model.

In Experiment 1, a higher number of cues led to a shift to more similarity-based strategies in judgment and categorization with more people following rule abstraction in the one-dimensional, linear task than in the multidimensional, linear task. Likewise, increasing the complexity of the functional relation made participants rely more upon similarity-based strategies. Dealing with a categorization problem (in comparison to a judgment problem), however, did not make participants shift more to similarity-based strategies. Moreover, more cues and a more complex functional relationship made it more difficult to predict which strategy people would apply in the second task given the strategy they applied in the task they solved first. To replicate these findings in a second experiment, we used a multidimensional quadratic task that is even closer to a function learning task in which individual preferences for rule-based and similarity-based learning should become more pronounced (McDaniel et al., 2013). As in Experiment 1, a categorization task did not change the amount of participants best described by a similarity-based strategy. Moreover, it was hard to predict which strategy people best described by a similarity-based strategy would rely on in the second task. However, people relying upon rules in the first task were more likely to shift to similarity-based processes in the second task. Taken together, these results suggest that providing scarce task feedback in categorization does not invite more similarity-based strategies per se. However, making rules more difficult or impossible to abstract not only triggers similarity-based strategies, but also harms the ability to consistently detect the strategy best suited to solve the task.

Memory Foundations of Human Judgment

Hoffmann, J A., von Helversen, B., & Rieskamp, J. (2013c). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. Manuscript submitted for publication.

The judgment and categorization literature particularly highlights that rule-based and similarity-based strategies may draw upon different knowledge representations (Hahn & Chater, 1998; Smith et al., 1998; but Pothos, 2005) and there has been a heated debate as to what degree these knowledge representations rest upon different memory abilities (Ashby & O'Brien, 2005; Knowlton, 1999; Lewandowsky, 2011; Newell, Dunn, & Kalish, 2011; Nosofsky & Zaki, 1998; Smith et al., 1998). Ashby and O'Brien (2005), for instance, suggested that executing simple rule-based categorization strategies requires working memory capacity, whereas exemplar retrieval involves episodic memory. In a similar vein, Juslin et al. (2008) argued that cue abstraction could be conceived as a capacity-constrained sequential process, whereas similarity-based judgment strategies might be driven by explicit or implicit memory. Although the role of working memory capacity for rule abstraction has earned a lot of attention in judgment and categorization showing that, for instance, learning even simple rules is impaired by working memory load (Filoteo et al., 2010; Zeithamova & Maddox, 2006), empirical evidence for the relationship between long-term memory and similarity-based strategies is still scarce (Ashby & O'Brien, 2005). Previous research has shown that exceptions to a rule, for instance, are recognized more often in a later recognition test (Davis, Love, & Preston, 2012; Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004). Likewise, the instruction to remember all exemplars by heart helps performance in judgment tasks that can only be solved by similarity-based strategies (Olsson et al., 2006). Dissociations between recognition and categorization performance between amnesic patients and healthy controls, in contrast, have been taken as evidence that similarity-based strategies may tap into both implicit and explicit long-term memory (Knowlton & Squire, 1993; Smith & Grossmann, 2008).

The third paper tried to shed some light on how memory abilities promote the selection and execution of rule-based and similarity-based judgment strategies and how these strategies, in turn, affect judgment performance. Specifically, we hypothesized that low

working memory capacity should hurt executing rule-based strategies, whereas difficulties with encoding and retrieval from episodic memory may harm similarity-based strategies. Moreover, working memory capacity may also facilitate discovering the appropriate judgment strategy, whereas episodic memory may only strengthen the preference for employing similarity-based strategies.

To investigate these questions, we conducted a study relating individual differences in memory abilities to judgment performance and judgment strategies in two different judgment tasks: A linear additive judgment task in which most participants should rely upon a rule-based judgment strategy and a multiplicative judgment task in which most participants should be best described by a similarity-based strategy (the same tasks as in Manuscript 2). Additionally, we measured working memory, episodic memory, and implicit memory by three different tests each. Classifying participants to the judgment strategies indeed confirmed that participants switched from a rule-based strategy in the linear judgment task to a similarity-based judgment strategy in the multiplicative task. To relate memory abilities to judgment performance we relied upon structural equation modeling. This analysis suggested that higher working memory capacity predicted higher judgment accuracy in linear judgment tasks, whereas the ability to solve multiplicative judgment tasks was predicted by episodic memory. Implicit memory was related to judgment performance neither in rule-based, nor in the similarity-based judgment tasks. Finally, better episodic memory also predicted choosing a similarity-based strategy in the multiplicative task and this choice of a similarity-based strategy enhanced judgment accuracy for similarity-based judgments. Working memory, in contrast, was linked to how well people executed the strategy learned in the linear judgment task and — ultimately — predicted judgment accuracy for rule-based judgments.

In sum, these results emphasize that not only task demands drive strategy shifts between rule-based and similarity-based processing, but judgment strategies also exploit different underlying cognitive abilities. While high working memory capacity may help

people to abstract rules, similarity-based strategies build upon the ability to encode and retrieve items from episodic memory. This suggests that focusing on cognitive abilities can help us to understand why people establish preferences for learning based upon rules or based upon exemplars.

General Discussion

Following a contingency approach to strategy selection, I outlined in the introduction that people may select a judgment strategy by trading off the accuracy a particular strategy can achieve with the effort necessary to execute this strategy. Within this framework, I focused on contrasting two types of judgment strategies: rule-based and similarity-based strategies. Replicating previous results, we found in all three manuscripts that the relative accuracy of rule-based and similarity-based strategies is one major determinant of strategy selection (Juslin et al., 2008; Karlsson et al., 2007). In Manuscript 1, we found a stronger switch to similarity-based strategies in a task in which reliance upon similarity is strongly enforced. Likewise, the functional relationship between cues and criterion fostered shifting to similarity-based strategies in judgment and categorization in the second paper. Indeed, even the same participants tended to rely more upon rules in linear tasks and more on similarity in multiplicative tasks (Manuscript 3).

Beyond accuracy, however, the effort that needs to be invested into strategy execution also affects which strategy people select. Increasing the difficulty to abstract rules —either by imposing a cognitive load on the decision maker (Manuscript 1) or by increasing the number of cues that need to be considered by a rule-based strategy (Manuscript 2) — enhanced reliance upon similarity-based strategies. These results dovetail research suggesting that providing knowledge about the cue directions (Platzer & Bröder, 2013; von Helversen et al., 2013) or a rule-based learning sequence (Juslin et al., 2008) facilitates abstraction of cue weights thereby fostering rule-based judgment strategies.

Finally, our third paper picked up the idea that people may also learn about the costs and benefits associated with each strategy and develop preferences over time for selecting one over another strategy. First, we found that the ability to solve rule-based and similarity-based judgment tasks hinges to a varying degree upon working memory and episodic memory. Furthermore, in line with research showing that older adults seem to avoid similarity-based strategies (Mata, von Helversen et al., 2012), we found that better episodic memory predicts how likely people are to select a similarity-based strategy over a rule-based strategy. Working memory capacity, in contrast, benefitted the ability to consistently execute learned strategies mimicking results suggesting that working memory is particularly important for executing learned rules (Del Missier et al., 2013).

In doing so, this dissertation establishes ties between different fields of psychology — from judgment to categorization to memory — showing how these fields can profit from the vast knowledge accumulated in each of those fields over time: Manuscript 2 reunified categorization with judgment research by investigating how the cognitive strategies underlying human judgment match strategies people follow to categorize objects. This manuscript showed that, indeed, the major task components leading to strategy shifts in categorization also encourage strategy shifts in judgment and vice versa. Manuscript 1 and 3 focused more on testing the memory representations underlying rule-based and similarity-based strategies yielding converging evidence that high working memory capacity may be involved to a larger extent in rule abstraction than in exemplar memory. In addition, manuscript 3 strongly reinforced the role of episodic memory for similarity-based judgments — a topic that has still received too little attention in categorization, judgment, and decision making.

In this dissertation, I offer contingency approaches as one conceptual framework to understand strategy selection. However, contingency approaches to strategy selection have not been left without critique: First, selecting a strategy may require applying a meta-strategy

to decide how to select the judgment strategy and hence simply move the strategy selection problem to a meta-level (Rieskamp & Otto, 2006). Second, although later approaches dropped the concept of a meta-strategy, these attempts to frame strategy selection as a function of effort and accuracy have been criticized as vague (Marewski & Schooler, 2011). To remedy these shortcomings, Rieskamp and Otto have suggested reinforcement learning as one mechanism helping to adapt the strategies to the task at hand. Alternatively, Marewski and Schooler proposed that the task environment, cognitive abilities, and the cognitive strategies mutually restrict the range of situations when a strategy can be applied. This dissertation supports the view that to advance our knowledge about strategy selection in judgment, categorization, and decision making, we need to consider in a common framework not only how people learn to adapt decision strategies to the task demands, but also how memory abilities may limit and shape the strategies we follow as routes to success.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481. doi:10.1037/0033-295X.105.3.442

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30*, 666–677.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences, 9*, 83–89. doi:10.1016/j.tics.2004.12.003

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning & motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *The Academy of Management Review, 3*, 439–449.

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition., 33*, 983–998. doi:10.1037/0278-7393.33.6.983

Börsenverein des deutschen Buchhandels (Eds.). (2013). *Buch und Buchhandel in Zahlen* [Book and bookselling trade in figures]. Frankfurt am Main, Germany: MVB Buchhändler-Vereinigung.

Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In B. Brehmer & C. R. B. Joyce (Eds.), *Human Judgment: The SJT view* (pp. 75–114). North-Holland: Elsevier.

Bröder, A. (2003). Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 29, 611–625. doi:10.1037/0278-7393.29.4.611

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*, 171-189. doi: 10.1006/jmps.1999.1282

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *The Quarterly Journal of Experimental Psychology, 65,* 439-464. doi: 10.1080/17470218.2011.608854

Cooksey, R. W. (1996). The methodology of social judgement theory. *Thinking & Reasoning, 2*, 141–174. doi: 10.1080/135467896394483

Davis, T., Love, B. C., & Preston, A. R. (2012). Learning the exception to the rule: model-based FMRI reveals specialized representations for surprising category members. *Cerebral Cortex, 22*, 260–273. doi:10.1093/cercor/bhr036

Del Missier, F., Mäntylä, T., Hansson, P., Bruine de Bruin, W., Parker, A. M., & Nilsson, L.-G. (2013). The multifold relationship between memory and decision making: An individual-differences study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1344-1364. doi: 10.1037/a0032379

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology, 60*, 1227–1245. doi:10.1080/17470210600926075

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review, 86*, 465–485. doi:10.1037//0033-295X.86.5.465

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107–140. doi:10.1037/0096-3445.127.2.107

Filoteo, J. V., Lauritzen, J. S., & Maddox, W. T. (2010). Removing the frontal lobes: The

    effects of engaging executive functions on perceptual category learning.

    Psychological Science, 21, 415–423. doi:10.1177/0956797610362646

Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Fast and frugal heuristics are

    plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas

    (2008). *Psychological Review, 115*, 230–237. doi:10.1037/0033-295X.115.1.230

Gigerenzer, G., Todd, P. M., & the ABC Research Group, (1999). Simple heuristics that

    make us smart. New York, NY, US: Oxford University Press.

Goldstein, D., & Gigerenzer, G. (2011). The beauty of simple models : Themes in

    recognition heuristic research. *Judgment and Decision Making, 6*, 392–395.

Graves, L., & Karren, R. (1992). Interviewer decision processes and effectiveness: An

    experimental policy-capturing investigation. *Personnel Psychology, 45*, 313–340.

    doi: 10.1111/j.1744-6570.1992.tb00852.x

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically

    distinguishable? *Cognition, 65*, 197–230. doi: 10.1016/S0010-0277(97)00044-9

Hoffmann, J A., von Helversen, B., & Rieskamp, J. (2013a). Deliberation's blindsight:

    How cognitive load can improve judgments. *Psychological Science, 24*, 869-879.

    doi: 10.1177/0956797612463581

Hoffmann, J A., von Helversen, B., & Rieskamp, J. (2013b). From rules to exemplars:

    Similar task features shape judgment and categorization processes. Manuscript to

    be submitted for publication.

Hoffmann, J A., von Helversen, B., & Rieskamp, J. (2013c). Pillars of judgment: How

    memory abilities affect performance in rule-based and exemplar-based judgments.

    Manuscript submitted for publication.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition, 106*, 259–298. doi:10.1016/j.cognition.2007.02.003

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132*, 133–156. doi:10.1037/0096-3445.132.1.133

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*, 404–426. doi:10.1037/0033-2909.134.3.404

Karlsson, L., Juslin, P., & Olsson, H. (2007). Adaptive changes between cue abstraction and exemplar memory in a multiple-cue judgment task with continuous cues. *Psychonomic Bulletin & Review, 14*, 1140–1146.

Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making : Contingent, not automatic, strategy shifts? *Judgment and Decision Making, 3*, 244–260.

Knowlton, B. J. (1999). What can neuropsychology tell us about category learning? *Trends in Cognitive Sciences, 3*, 123–124. doi: 10.1016/S1364-6613(99)01292-9

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262*, 1747–1749. doi: 10.1126/science.8259522

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124*, 83–97. doi:10.1037//0096-3445.124.1.83

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 720-738. doi: 10.1037/a0022639

Lewandowsky, S., Yang, L. X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*, 881-904. doi: 10.1037/a0027298

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes, 66*, 309–332. doi:10.1016/j.beproc.2004.03.011

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review, 118*, 393–437. doi:10.1037/a0024143

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science, 17*, 944–948. doi:10.1111/j.1467-9280.2006.01809.x

Mata, R., Pachur, T., von Helversen, B., Hertwig, R., Rieskamp, J., & Schooler, L. (2012). Ecological rationality: a framework for understanding and aiding the aging decision maker. *Frontiers in Neuroscience, 6*, 1-19. doi: 10.3389/fnins.2012.00019

Mata, R., von Helversen, B., Karlsson, L., & Cüpper, L. (2012). Adult age differences in categorization and multiple-cue judgment. *Developmental Psychology, 48*, 1188–1201. doi:10.1037/a0026084

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2013). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General*. Advance online publication. doi:10.1037/a0032963

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*, 10–17.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238. doi:10.1037/0033-295X.85.3.207

Newell, B., Dunn, J., & Kalish, M. (2011). Systems of category learning: fact or fantasy?
In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 167-215). Burlington: Academic Press.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 700–708. doi: 10.1037/0278-7393.14.4.700

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53–79. doi:10.1037/0033-295X.101.1.53

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255. doi: 10.1111/1467-9280.00051

Olsson, A.-C., Enkvist, T., & Juslin, P. (2006). Go with the flow: How to master a nonlinear multiple-cue judgment task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1371–1384. doi: 10.1037/0278-7393.32.6.1371

Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology, 65*, 207–240. doi:10.1016/j.cogpsych.2012.03.003

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 548–568. doi: 10.1037/0278-7393.21.3.548

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY, US: Cambridge University Press.

Platzer, C., & Bröder, A. (2013). When the rule is ruled out: Exemplars and rules in decisions from memory. *Journal of Behavioral Decision Making, 26,* 429-441. doi: 10.1002/bdm.1776

Pothos, E. M. (2005). The rules versus similarity distinction. *The Behavioral and Brain Sciences, 28*, 1–14; discussion 14–49.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135*, 207–236. doi:10.1037/0096-3445.135.2.207

Rothstein, H. G. (1986). The effects of time pressure on judgment in multiple cue probability learning. *Organizational Behavior and Human Decision Processes, 37*, 83–92. doi:10.1016/0749-5978(86)90045-2

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General, 133*, 534–553. doi:10.1037/0096-3445.133.4.534

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22. doi:10.1037/0033-2909.119.1.3

Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience and Biobehavioral Reviews, 32*, 249–264. doi:10.1016/j.neubiorev.2007.07.009

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65*, 167–196.

von Helversen, B., Karlsson, L., Mata, R., & Wilke, A. (2013). Why does cue polarity information provide benefits in inference problems? The role of strategy selection

and knowledge of cue importance. *Acta Psychologica, 144*, 73–82.

doi:10.1016/j.actpsy.2013.05.007

von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue abstraction processes in multiple-cue judgment. *Developmental Psychology, 46*, 220–229. doi:10.1037/a0016690

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General, 137*, 73–96. doi:10.1037/0096-3445.137.1.73

Von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-based and exemplar-based processes compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 867–889. doi:10.1037/a0015501

Wigton, R. S. (1996). Social judgement theory and medical judgement. *Thinking and Reasoning, 2*, 175-190. doi: 10.1080/135467896394492

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology, 59*, 555–561.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition, 34*, 387–398. doi:10.3758/BF03193416

Deliberation's Blindsight: How Cognitive Load Can Improve Judgments

Janina A. Hoffmann, Bettina von Helversen, and Jörg Rieskamp
University of Basel

Abstract

Multitasking poses a major challenge in modern work environments by putting the worker under cognitive load. Performance decrements often occur when people are under high cognitive load because they switch to less demanding—and often less accurate—cognitive strategies. Although cognitive load disturbs performance over a wide range of tasks, it may also carry benefits. In the experiments reported here, we showed that judgment performance can increase under cognitive load. Participants solved a multiple-cue judgment task in which high performance could be achieved by using a similarity-based judgment strategy but not by using a more demanding rule-based judgment strategy. Accordingly, cognitive load induced a shift to a similarity-based judgment strategy, which consequently led to more accurate judgments. By contrast, shifting to a similarity-based strategy harmed judgments in a task best solved by using a rule-based strategy. These results show how important it is to consider the cognitive strategies people rely on to understand how people perform in demanding work environments.

Suppose you hurt your leg in an accident and go to the hospital for emergency treatment. While treating you, the physician is repeatedly interrupted by a medical assistant. Is the physician still able to treat you properly? Emergency physicians are—on average—interrupted 10 times per hour (Chisholm, Collison, Nelson, & Cordell, 2000). These interruptions can increase the risk of failure, such as medication errors (Westbrook, Woods, Rob, Dunsmuir, & Day, 2010). One reason why distractions are so damaging is that they increase cognitive load on the physician and reduce working memory capacity for the focal task (Baddeley, 1992; Baddeley & Hitch, 1974).

Research has shown that high cognitive load severely impairs performance in various tasks, ranging from memory (Baddeley & Hitch, 1974) to motor abilities (Yogev-Seligmann, Hausdorff, & Giladi, 2008) to problem solving (Logie, Gilhooly, & Wynn, 1994). Similarly, making accurate judgments, such as diagnosing a patient, can require high working memory capacity, and thus accuracy should suffer under cognitive load (Juslin, Karlsson, & Olsson, 2008; Payne, Bettman, & Johnson, 1993; Weaver & Stewart,

2012). Sometimes, however, cognitive load can improve performance: For instance, experienced golf players who are distracted putt better than experienced golf players focusing on performance aspects (Beilock, Carr, MacMahon, & Starkes, 2002). Likewise, cognitive load induced by the presence of other people often facilitates performance (e.g., Baron, 1986; Markman, Maddox, & Worthy, 2006). Given that negative consequences of cognitive load are often, but not always, found, under what circumstances does performance increase under cognitive load?

To predict performance, we argue that one must consider the cognitive strategies people use for solving problems and how well these strategies perform. Research shows that strategies demanding high working memory capacity are impaired under cognitive load, which induces

Corresponding Author:
Janina A. Hoffmann, Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland
E-mail: janina.hoffmann@unibas.ch

people to switch to less demanding strategies (Beach & Mitchell, 1978; Beilock & DeCaro, 2007; Payne et al., 1993; Rieskamp & Hoffrage, 2008). If less demanding strategies cannot help solve the task, performance decreases. However, if less demanding strategies can help solve the task, performance can increase (Beilock & DeCaro, 2007). Social pressure, for instance, expedites learning in non-verbalizable categorization problems (Markman et al., 2006) that are solvable by using similarity-based strategies (Juslin, Olsson, & Olsson, 2003) but harms learning in verbalizable categorization problems solvable by using rule-based strategies.

In the present work, we investigated how cognitive load changes strategy use in a multiple-cue judgment task and how strategy use interacts with the task environment. Specifically, we first tested whether cognitive load fosters switching from a rule-based judgment strategy to a similarity-based judgment strategy. Second, we tested whether cognitive load improves performance in tasks for which the similarity-based strategy is better suited.

Multiple-Cue Judgments

In multiple-cue judgment tasks, a number of cues, such as a patient's symptoms, are used to predict a quantitative criterion, say, an appropriate drug dosage for that patient. Recent research suggests that people commonly use two types of cognitive strategies for judgments: rule-based strategies and similarity-based strategies (Erickson & Kruschke, 1998; Juslin et al., 2008; Nosofsky, Palmeri, & McKinley, 1994; von Helversen & Rieskamp, 2008, 2009). Rule-based strategies assume that people try to find or abstract a rule specifying the relation between each cue and the criterion. The abstracted cue weights are then integrated in a linear additive fashion. For instance, a physician may apply a rule that specifies the appropriate dosage as a linear function of the patient's symptoms. Linear regression models can capture these rules and have successfully described human judgment in various domains (Brehmer & Brehmer, 1988).

Alternatively, physicians could recall patients they have previously treated and estimate the dosage according to the treatment of similar patients. In this case, the physician relies on a similarity-based strategy. Models assuming a similarity-based strategy, such as exemplar models, successfully predict human behavior in a wide selection of cognitive tasks, such as categorization (Juslin et al., 2003; Nosofsky & Johansen, 2000) and judgment (Juslin et al., 2008). Exemplar models assume that previously encountered exemplars are stored in memory. When judging a new object, the similarity of this "probe" to all stored objects determines the judgment (see Section A

in the Supplemental Material available online for the models' mathematical descriptions).

Converging evidence suggests that people switch between rule- and similarity-based strategies depending on task characteristics (Juslin et al., 2003; Juslin et al., 2008; von Helversen, Mata, & Olsson, 2010). For instance, Juslin and colleagues (2008) found that people used a rule-based cue-abstraction strategy in a linear judgment problem in which the criterion was an additive function of the cues. However, people switched to an exemplar strategy in a nonlinear task in which the criterion was a multiplicative function of the cues. Likewise, cognitive load may induce selecting another judgment strategy. In fact, evidence suggests that rule-based strategies demand more working memory capacity than similarity-based strategies (Juslin et al., 2008). For instance, increased cognitive load impaired performance in rule-based categorizations but marginally affected performance in similarity-based categorizations (Zeithamova & Maddox, 2006, 2007; but see Miles & Minda, 2011). Furthermore, Filoteo, Lauritzen, and Maddox (2010) found that cognitive load improved performance in similarity-based, but not in rule-based, categorizations; they explained that this improvement occurred because more people shifted to implicit procedural strategies when making similarity-based categorizations. Sloman (1996) argued that similarity-based processes are executed automatically and require little working memory capacity. However, to what extent similarity-based strategies draw on working memory is still debated (Ashby & O'Brien, 2005; Juslin et al., 2008; Karlsson, Juslin, & Olsson, 2008; Lewandowsky, 2011; Nosofsky & Zaki, 1998).

Following this debate, we investigated how cognitive load affects judgment strategies and performance. If working memory limitations affect rule-based strategies more than similarity-based strategies, increased cognitive load should promote a shift from rule-based to similarity-based judgments. Furthermore, when similarity-based strategies are better suited for solving the judgment problem—as in nonlinear judgment tasks—cognitive load may even enhance performance.

Study 1: Cognitive Load in a Nonlinear Judgment Task

To test our hypothesis, we trained participants in the present study to predict the criterion value for a number of objects using five cues. The criterion was a nonlinear, multiplicative function of the cues and could be better predicted by a similarity-based strategy than by a rule-based strategy (von Helversen & Rieskamp, 2008). We manipulated cognitive load with a concurrent memory task in three conditions that differed according to

whether participants were given no, low, or high cognitive load.

Method

Participants. Ninety participants (42 women, 48 men; mean age = 24 years, *SD* = 5 years) were recruited from the University of Basel. Participants received 17 Swiss francs (CHF) per hour (roughly $18) and a performance-contingent bonus (*M* = 8.3 CHF) for participation. One participant who always made identical judgments was excluded from the analysis.

Design and materials. The cover story in the judgment task was adopted from von Helversen et al. (2010) and asked participants to predict how many fictitious creatures ("Golbis") a comic figure (a "Sonic") could catch. The Sonics' appearance differed in five binary features (the cues): hair (spiky vs. dreadlocks), nose (red round vs. yellow beaky), tail (spiny vs. curly), ears (pointy vs. floppy), and body (green wings vs. blue spikes).[1] These cues could be used to predict how many Golbis a Sonic would catch (the criterion). Table 1 illustrates the task structure: The cues were given a binary value of zero or one, and they varied in their cue weights, that is, in their importance for predicting the criterion. The cue weights were randomly assigned to the five pictorial cues, as were the cue values (zero or one) to the features (e.g., spiny vs. curly). We divided the items into a training set and a validation set; both sets could be better solved

Table 1. Cue and Criterion Values of Items in the Nonlinear Judgment Task of Study 1

Set and item	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5	Criterion
Training set						
Item 1	1	1	0	1	1	20
Item 2	0	0	0	1	0	1
Item 3	0	0	0	0	0	0
Item 4	0	1	0	0	1	2
Item 5	1	1	0	0	1	7
Item 6	1	0	0	1	1	5
Item 7	0	0	0	0	1	0
Item 8	1	1	0	1	0	9
Item 9	0	1	0	0	0	1
Item 10	0	0	1	0	0	1
Item 11	0	1	0	1	0	2
Item 12	0	1	0	1	1	5
Item 13	0	0	1	1	1	4
Item 14	1	0	1	1	1	16
Item 15	1	1	0	0	0	3
Item 16	1	1	1	1	1	62
Validation set						
Item 1	1	0	0	1	0	2
Item 2	0	0	1	1	0	2
Item 3	0	1	1	0	0	2
Item 4	1	0	0	0	0	1
Item 5	1	0	1	1	0	7
Item 6	1	0	1	0	1	6
Item 7	1	1	1	0	1	23
Item 8	1	0	1	0	0	3
Item 9	0	0	1	0	1	1
Item 10	0	1	1	1	0	6
Item 11	0	0	0	1	1	1
Item 12	0	1	1	0	1	5
Item 13	0	1	1	1	1	14
Item 14	1	0	0	0	1	2
Item 15	1	1	1	1	0	28
Item 16	1	1	1	0	0	10

Note: Training items were presented in the training and the test phase. Validation items appeared only during the test phase.

by a similarity-based strategy than a rule-based strategy. Additionally, the two strategies predicted different responses on the validation items (for item selection, see Section A in the Supplemental Material).

Procedure. To control for possible differences in working memory capacity, we first asked participants to complete an operation-span task (Unsworth, Heitz, Schrock, & Engle, 2005). During this task, participants recalled letters while solving mathematical equations. The subsequent judgment task was divided into a training and a test phase. In the training phase, participants learned to make judgments for 16 training items. To induce a shift to a similarity-based strategy, we manipulated cognitive load during this phase across three conditions, which differed according to whether participants were given no, low, or high cognitive load. Thirty participants were assigned to each condition.

On each trial in the training phase, participants saw 1 of the 16 Sonics from the training set and estimated its criterion value. After each trial, participants received feedback about the correct criterion value and the points earned. In the low- or high-cognitive-load condition, participants saw two or four consonants, respectively, before the Sonic appeared. Consonants were presented consecutively, each for 2 s. After the participants received feedback about their criterion judgment, they were asked to recall the letters in their presentation order. The training phase ended when a learning criterion was reached. Participants met this learning criterion when judgment accuracy, as measured in root-mean-square deviation (RMSD) between participants' judgments and the criterion values, fell below 6 RMSD. Each participant completed at least 8 training blocks, each consisting of 16 trials; training terminated after 14 blocks even if the learning criterion had not been reached. In the test phase,

participants estimated criterion values for all 32 Sonics from the training and the validation sets twice without feedback and without cognitive load.

To motivate participants, we provided a performance-contingent payment. In each trial, participants earned 10 points (corresponding to 0.05 CHF) for a correct answer. The more their judgment deviated from the correct answer, the fewer points they received: They received 9 points if their judgment deviated by one from the correct answer, 8 points if it deviated by two, 6 points if it deviated by three, and 3 points if it deviated by four. Participants under low and high cognitive load received an additional point for correct letter recall. To prevent participants from trading off recall performance and judgment performance, we did not award any points for the memory or for the judgment task when they could not recall the letters. Additionally, participants received a bonus of 3 CHF if they reached the learning criterion for the judgment task within 14 training blocks.

Results

Adherence to cognitive load. To check whether we manipulated cognitive load successfully, we calculated the percentage of correctly recalled letter sequences over all blocks. Letter recall was generally high; however, participants under low cognitive load recalled letters better than did participants under high cognitive load, $t(46.13) = 3.35$, $p = .002$ (see Table 2). In both conditions, higher criterion-judgment accuracy was related to better letter recall (all rs < −.35, all ps < .05), which indicates that participants did not trade off letter recall and judgment accuracy. Excluding participants who recalled fewer than 90% of the letter sequences correctly led to comparable results. Taken together, these results suggest that the cognitive-load manipulation was successful.

Table 2. Mean Results for the Three Conditions in Study 1

Phase and measure	No cognitive load	Low cognitive load	High cognitive load
Pretraining phase			
Operation-span score	37.5 (16.3)	36.0 (17.2)	42.7 (19.2)
Training phase			
Letters recalled (%)	—	96.0 (4.3)	90.7 (7.5)
Number of blocks completed	11.5 (2.5)	10.0 (2.3)	10.8 (2.8)
Judgment accuracy: last block	8.14 (5.63)	7.40 (6.62)	8.54 (6.52)
Test phase			
Judgment accuracy: training set	8.03 (3.95)	8.79 (5.46)	10.49 (6.57)
Judgment accuracy: validation set	12.87 (6.43)	10.55 (4.54)	9.30 (3.47)
Judgment accuracy: both sets	11.21 (4.20)	10.20 (3.91)	10.30 (4.41)

Note: Standard deviations are given in parentheses. Judgment accuracy was measured in root-mean-square deviations (RMSD) from the correct response.

Differences in working memory capacity. Working memory capacity may be an important mediator of judgment performance (DeCaro, Carlson, Thomas, & Beilock, 2009; Lewandowsky, 2011). Hence, we measured individual differences in working memory capacity with an operation-span task (Unsworth et al., 2005). Working memory capacity did not vary significantly between the cognitive-load conditions, $F(2, 86) = 1.20$, $p = .305$ (see Table 2). Including working memory capacity as a covariate did not affect the results in any subsequent analysis.

Criterion-judgment performance. Can people learn accurate judgments even under high cognitive load? The majority of participants (68%) reached the learning criterion within 14 blocks, which suggests that, overall, participants mastered the task. The number of participants who did not reach the learning criterion did not differ significantly among conditions (high load: 11, low load: 6, control: 12), $\chi^2(2, N = 89) = 2.85$, $p = .241$.[2] Additionally, we assessed learning performance with the number of training blocks completed and judgment accuracy in the last training block (see Table 2). An analysis of variance revealed that participants in the two cognitive-load conditions did not require more training than participants without cognitive load, $F(2, 86) = 2.30$, $\eta^2 = .05$, $p = .107$. Neither high nor low cognitive load resulted in poorer judgment accuracy in the last training block than did no cognitive load, $F(2, 86) < 1$, $p = .778$. These results show that cognitive load did not harm learning.

But were participants able to generalize the good performance to validation items when they learned under cognitive load? We measured judgment accuracy in the test phase as the RMSD between the criterion value and participants' judgments, averaged over the two test blocks separately for training and validation items. As expected based on the learning results, performance for training items did not differ significantly among the three conditions, $F(2, 86) = 1.61$, $\eta^2 = .04$, $p = .206$ (see Table 2). However, for validation items, participants made more accurate judgments in the two cognitive-load conditions than in the no-load condition, $F(2, 86) = 4.00$, $\eta^2 = .09$, $p = .022$. Furthermore, a linear contrast for cognitive load showed that for validation items, increasing cognitive load led to higher judgment accuracy, $F(2, 86) = 7.78$, $p = .007$. In sum, consistent with our hypothesis, the results showed that cognitive load increased people's judgment performance.

Cognitive modeling of judgment strategies. According to our hypothesis, cognitive load might induce people to switch to a similarity-based strategy. Because the task could be better solved with a similarity-based strategy than with a rule-based strategy, such a shift could explain performance improvements under cognitive load.

We followed a cognitive-modeling approach to investigate the judgment strategies participants used. We first fitted three computational models, an exemplar model (similarity-based strategy), a linear model (rule-based strategy), and a baseline model (estimating participants' mean judgment), to participants' judgments during the training phase (for details, see Section A in the Supplemental Material). We then determined how accurately the models predicted participants' judgments during the test phase and excluded participants best described by the baseline model. To capture how much participants relied on a linear versus an exemplar model, we fitted a strategy weight ranging from zero to one to participants' judgments in the test phase. This strategy weight weighs the predictions of the linear and the exemplar model for the test phase. A strategy weight over .5 indicates a higher probability for the exemplar model; a strategy weight lower than .5 indicates a higher probability for the linear model. Classifying participants based on a threshold strategy weight of .5 was identical to a classification based on model fit in the test phase.

Cognitive load, indeed, affected the strategy weight, $F(2, 67) = 6.98$, $\eta^2 = .17$, $p = .005$. Participants under high cognitive load had a higher strategy weight ($M = .86$, $SE = .04$) than did participants under low cognitive load ($M = .70$, $SE = .07$) or without cognitive load ($M = .52$, $SE = .07$). Figure 1 (upper panel) illustrates the effect of cognitive load on strategy use, with participants classified based on the strategy weight. In the control condition, the linear and the exemplar model predicted an equal percentage of participants best. However, under cognitive load, the exemplar model predicted the majority of participants best. In addition to cognitive load, working memory capacity may alter strategy choice. To analyze this relationship, we regressed strategy weight on working memory capacity using cognitive load as a covariate. In this analysis, working memory capacity did not predict strategy weight beyond cognitive load, $b = 0.001$, $SE = 0.002$, $t(67) = 0.54$, $p = .592$. Taken together, these results suggest that cognitive load induced participants to rely more on a similarity-based than a rule-based judgment strategy.

Judgment accuracy and cognitive models. Can a change of strategy explain differences in judgment accuracy under cognitive load? Figure 1 (lower panel) shows judgment accuracy for validation items in the test phase, separately for participants assigned to the exemplar and the linear model. The figure illustrates that participants assigned to the exemplar model judged validation items more accurately than did participants assigned to the linear model.

If cognitive load increases judgment performance by changing the cognitive strategy, the strategy weight should mediate the effect of cognitive load on judgment
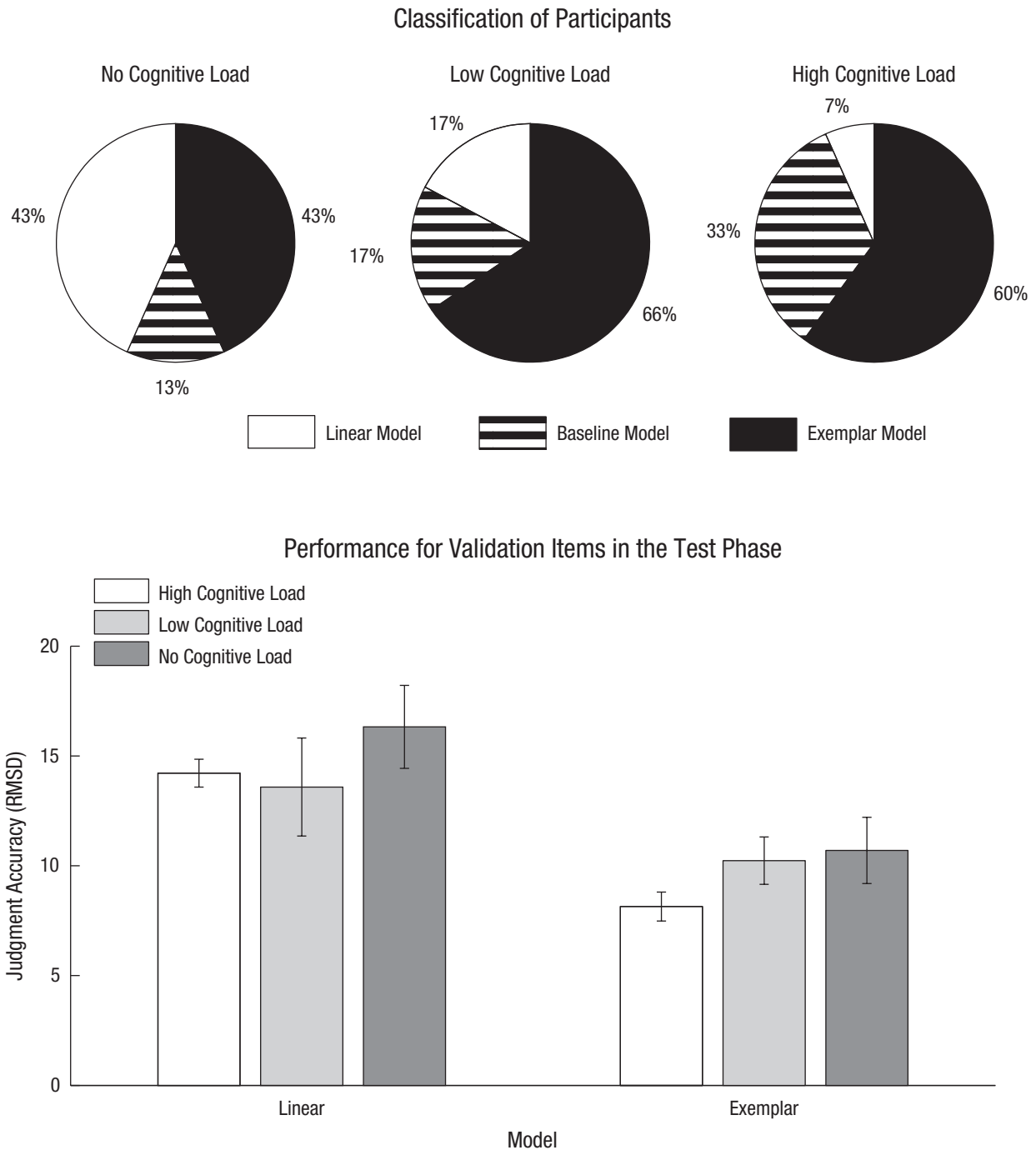
Classification of Participants

Fig. 1. Judgment strategies in the nonlinear judgment task in Study 1. The pie charts show the percentage of participants in each of the three cognitive-load conditions who were best described by the baseline, the linear, or the exemplar model. The graph shows judgment accuracy, measured in root-mean-square deviations (RMSDs) from the correct response, for validation items in the test phase as a function of model type and cognitive-load condition. Error bars represent ±1 *SE*.

performance. We tested this hypothesis with a mediation analysis in which cognitive load was the independent variable, strategy weight was the mediator, and judgment accuracy for validation items was the dependent variable (Preacher & Hayes, 2008). First, we regressed judgment accuracy on cognitive load. This regression showed that increasing cognitive load led to a higher judgment accuracy, $b = -2.39$, $SE = 0.78$, $t(68) = 3.08$, $p = .003$, $R^2 = .12$. However, with strategy weight included in the hierarchical regression, cognitive load no longer predicted

judgment accuracy, $b = -0.95$, $SE = 0.75$, $t(67) = 1.27$, $p = .209$. Instead, the strategy weight predicted judgment accuracy, $b = -8.35$, $SE = 1.79$, $t(67) = 4.65$, $p < .001$, $R^2 = .34$. A test of the indirect effect indicated that the strategy weight mediated the effect of cognitive load on judgment, $b = -1.44$, $SE = 0.49$, Sobel's $Z = 2.96$, $p = .003$, and thus explains why participants performed better under cognitive load (see Section A in the Supplemental Material for additional results).

Study 2: Extension to a Linear Judgment Task

How does cognitive load influence performance in a linear task? In a linear judgment task, similarity-based strategies lead to worse performance than rule-based strategies. Thus, if cognitive load causes participants to rely more on a less demanding similarity-based strategy than a more demanding rule-based strategy, this should decrease performance in a linear task. However, strategy selection is affected not only by the effort it takes to process a strategy, but also by feedback about strategy performance (Payne et al., 1993). Feedback reinforces successful strategies and makes their selection more likely (Rieskamp & Otto, 2006). In the nonlinear task, feedback and cognitive load promoted reliance on similarity-based strategies. Yet, in a linear judgment task, feedback should favor a rule-based strategy. Accordingly, participants may be more motivated to use a rule-based strategy, which would reduce the influence of cognitive load on strategy selection. To investigate this question, we compared how people under high cognitive load (four letters) and people without cognitive load solved a linear judgment task (see Mata, von Helversen, Karlsson, & Cüpper, 2012).

Method

Sixty participants (35 women, 25 men; mean age = 25 years, SD = 7 years) solved a linear judgment task. Each participant was randomly assigned to one of two conditions: high cognitive load (in which participants saw four letters before each trial, as in Study 1) or no cognitive load (in which participants saw no letters before each trial). Participants received 17 CHF per hour and a performance-contingent bonus (M = 5.4 CHF). The design and materials were the same as in Study 1, except that the Sonics' appearance varied among only four binary cues: hair, nose, ears, and body. The criterion was a linear function of these four cues. The task consisted of a training and a test phase. During the training phase, participants repeatedly judged 10 training items until a learning criterion had been reached (with at least 8 and at most 16 blocks). In the test phase, participants judged 10 training items and 6 validation items four times without feedback (see Section B in the Supplemental Material for details on the methods used in Study 2).

Results

To test whether people switched to a similarity-based strategy in the present study, we followed the same approach as in the first study. We modeled participants' judgments and excluded participants assigned to the baseline model. Then we estimated the strategy weight to capture how much participants relied on an exemplar rather than a linear model. As illustrated in Figure 2 (upper panel), the percentage of participants assigned to the exemplar model increased slightly under cognitive load, reflected in a marginally significant higher strategy weight in the high-load condition (M = .53, SE = .07) than in the control condition, (M = .33, SE = .07), $t(49) = 1.91$, $d = 0.54$, $p = .061$.

Cognitive load did not affect performance (high-load condition: M = 2.32 RMSD, SD = 0.97; control condition: M = 2.15 RMSD, SD = 0.93), $t(58) = 0.68$, $p = .50$. A regression analysis, however, showed that a higher strategy weight representing similarity-based strategies predicted lower judgment performance on validation items, $b = -0.873$, $SE = 0.319$, $t(49) = 2.735$, $p = .009$, $R^2 = .13$. Thus, a similarity-based strategy harmed judgment performance in the linear task (see Fig. 2, lower panel).

In sum, cognitive load induced a shift to similarity-based strategies even in a linear judgment task. Furthermore, following a similarity-based strategy harmed judgment performance. However, the shift was not pronounced enough to effectively decrease performance under high cognitive load (see Section B in the Supplemental Material for a more detailed analysis of results of Study 2).

Discussion

In daily life, gaining time by doing several things at once is tempting. Although most people can walk and talk at the same time, using a mobile phone while driving can be dangerous. In fact, distraction impairs performance over a wide range of tasks (Baddeley, 1992). Distractions, however, may not always hurt performance. In contrast, we found that people made more accurate judgments after learning a nonlinear judgment task under concurrent memory load, a finding that matches research showing that cognitive load can enhance performance (Beilock & DeCaro, 2007; Filoteo et al., 2010; Markman et al., 2006).

In our research, we extended these findings to judgments by modeling the cognitive strategies people use and linking these strategies to judgment performance. In
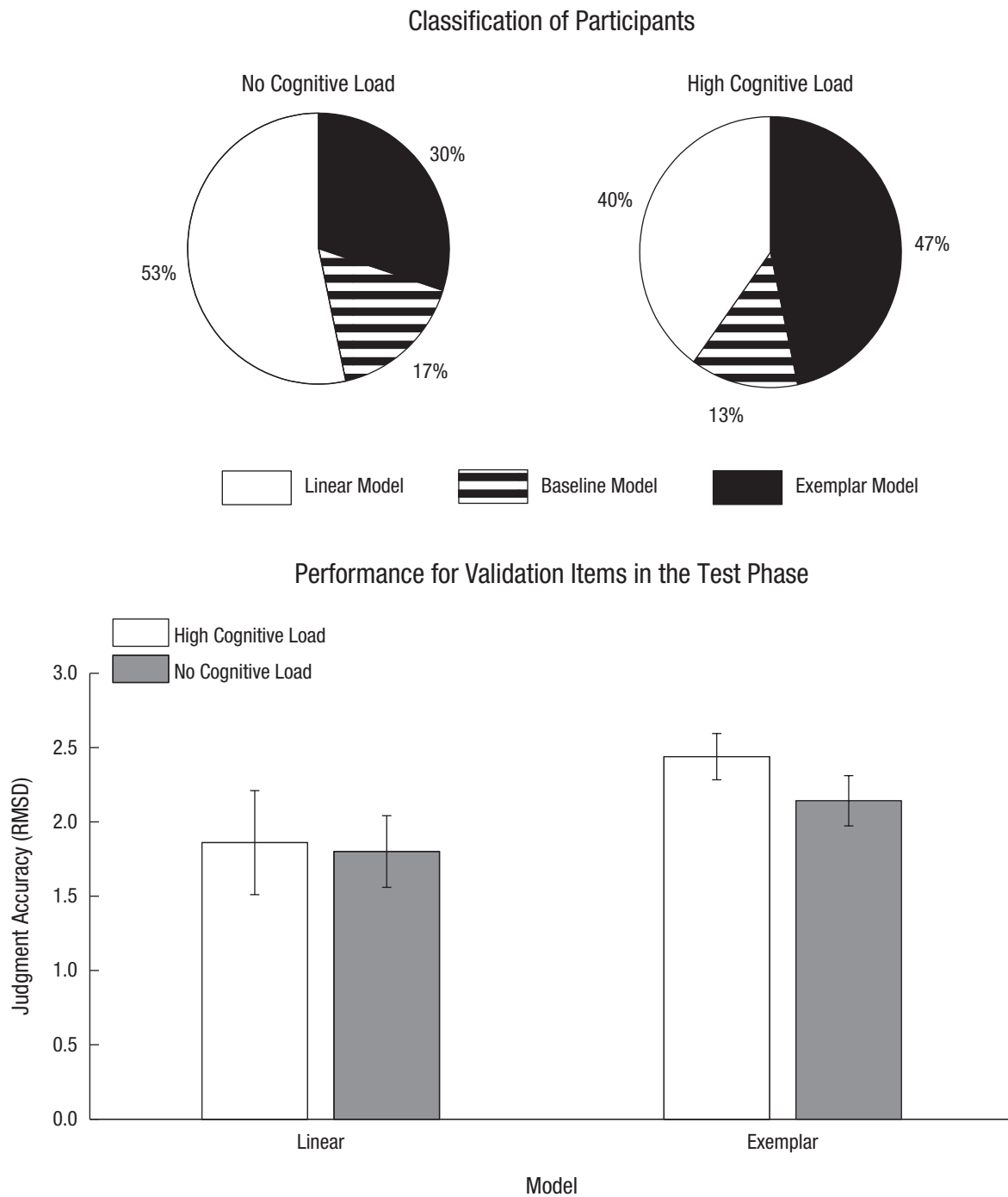
Classification of Participants

No Cognitive Load   High Cognitive Load

Linear Model   Baseline Model   Exemplar Model

Performance for Validation Items in the Test Phase

Fig. 2. Judgment strategies in the linear judgment task in Study 2. The pie charts show the percentage of participants in the two cognitive-load conditions who were best described by the baseline, the linear, or the exemplar model. The graph shows judgment accuracy, measured in root-mean-square deviations (RMSDs) from the correct response, for validation items in the test phase as a function of model type and cognitive-load condition. Error bars represent ±1 *SE*.

the nonlinear judgment task, cognitive load increased performance for validation items. This performance increase was explained by a shift from a rule-based strategy to a less demanding but more accurate similarity-based strategy. Switching to a less demanding strategy,

however, does not always benefit judgment performance. If the strategy people use under cognitive load is not adapted to the judgment problem, judgment performance can decrease. Accordingly, in a linear judgment task, we found that following the less accurate similarity-based

strategy impaired judgment performance. This suggests that considering the cognitive strategies people use under cognitive load is crucial for predicting performance.

Our results resonate with research suggesting that cognitive load induces people to switch to a less demanding cognitive strategy (Beach & Mitchell, 1978; Beilock & DeCaro, 2007; Payne et al., 1993; Rieskamp & Hoffrage, 2008). In the two experiments reported here, we found that participants under cognitive load were more likely to use a similarity-based strategy than participants who were not under cognitive load. One reason for this strategy change could be that rule-based strategies are more susceptible to working memory limitations than similarity-based strategies are (Juslin et al., 2008). This is supported by research suggesting that rule-based strategies place strong demands on working memory (Ashby & O'Brien, 2005; Zeithamova & Maddox, 2006, 2007), whereas similarity-based categorization may be learned via implicit, automatic processes (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & O'Brien, 2005; Filoteo et al., 2010; Markman et al., 2006, but see Karlsson et al., 2008; Lewandowsky, 2011).

The tasks, however, differed in how much participants shifted their strategies under cognitive load. In the linear judgment task, participants relied less strongly on a similarity-based strategy than participants did in the nonlinear task. Possibly, performance feedback reinforced rule-based strategies enough to motivate participants to rely on a rule-based strategy that allowed accurate judgments to outweigh effort reductions from switching to a similarity-based strategy (Payne et al., 1993).

The effect of cognitive load may also depend on type of load: In our studies, we focused on verbal cognitive load. Visual load, however, interferes more strongly with visual processing and reduces learning in similarity-based categorizations (Miles & Minda, 2011). Thus, high visual cognitive load may impair similarity-based judgment strategies. Additionally, the effect of cognitive load may depend on reward structure (Maddox & Markman, 2010; Worthy, Markman, & Maddox, 2009). Under high pressure, aiming to minimize losses impairs performance in similarity-based categorizations (Worthy et al., 2009). In our studies, participants tried to maximize gains by collecting as many points as possible. Yet it is possible that avoiding losses would hurt similarity-based judgments under cognitive load.

In sum, we found that people under cognitive load relied more often on a similarity-based judgment strategy than on a rule-based judgment strategy. Although this strategy change proved useful in a nonlinear judgment task, following a similarity-based strategy harmed performance in a linear judgment task. Evidently, recognizing the cognitive strategies that people rely on is a key to understanding how they solve problems and can help researchers predict how good people are at solving them. Uncovering people's cognitive strategies may lead to a better understanding of when and how people can maintain high performance even in distracting environments, such as emergency departments.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Supplemental Material

Additional supporting information may be found at http://pss .sagepub.com/content/by/supplemental-data

Notes

1. We used one cue more in the present study than von Helversen, Mata, and Olsson (2010) did. To make sure that the additional cue (the tail) was as salient as the other cues, we asked 10 participants to name the differences among the most dissimilar Sonics.
2. Including only participants who learned the task yielded the same conclusions as the analysis based on the complete data set.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481. doi:10.1037/0033-295X.105.3.442

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, *9*, 83–89. doi:10.1016/j.tics.2004.12.003

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559. doi:10.1126/science.1736359

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning & motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press. doi:10.1016/ S0079-7421(08)60452-1

Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 1–40). New York, NY: Academic Press.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *The Academy of Management Review*, *3*, 439–449. doi:10.5465/AMR.1978 .4305717

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied*, *8*, 6–16. doi:10.1037//1076-898X.8.1.6

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 983–998. doi:10.1037/0278-7393.33.6.983

Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75–114). Amsterdam, The Netherlands: Elsevier.

Chisholm, C. D., Collison, E. K., Nelson, D. R., & Cordell, W. H. (2000). Emergency department workplace interruptions: Are emergency physicians "interrupt-driven" and "multitasking"? *Academic Emergency Medicine*, *7*, 1239–1243. doi:10.1111/j.1553-2712.2000.tb00469.x

DeCaro, M. S., Carlson, K. D., Thomas, R. D., & Beilock, S. L. (2009). When and how less is more: Reply to Tharp and Pickering. *Cognition*, *111*, 415–421. doi:10.1016/j.cognition.2009.03.001

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107–140. doi:10.1037/0096-3445.127.2.107

Filoteo, J. V., Lauritzen, J. S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science*, *21*, 415–423. doi:10.1177/0956797610362646

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*, 259–298. doi:10.1016/j.cognition.2007.02.003

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*, 133–156. doi:10.1037/0096-3445.132.1.133

Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts? *Judgment and Decision Making*, *3*, 244–260.

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 720–738. doi:10.1037/a0022639

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition*, *22*, 395–410. doi:10.3758/BF03200866

Maddox, W. T., & Markman, A. B. (2010). The motivation-cognition interface in learning and decision-making. *Current Directions in Psychological Science*, *19*, 106–110. doi:10.1177/0963721410364008

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, *17*, 944–948. doi:10.1111/j.1467-9280.2006.01809.x

Mata, R., von Helversen, B., Karlsson, L., & Cüpper, L. (2012). Adult age differences in categorization and multiple-cue judgment. *Developmental Psychology*, *48*, 1188–1201. doi:10.1037/a0026084

Miles, S. J., & Minda, J. P. (2011). The effects of concurrent verbal and visual tasks on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 588–607. doi:10.1037/a0022309

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375–402.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53–79. doi:10.1037/0033-295X.101.1.53

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*, 247–255. doi:10.1111/1467-9280.00051

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891. doi:10.3758/BRM.40.3.879

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, *127*, 258–276. doi:10.1016/j.actpsy.2007.05.004

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*, 207–236. doi:10.1037/0096-3445.135.2.207

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22. doi:10.1037/0033-2909.119.1.3

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. doi:10.3758/BF03192720

von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue abstraction processes in multiple-cue judgment. *Developmental Psychology*, *46*, 220–229. doi:10.1037/a0016690

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General*, *137*, 73–96. doi:10.1037/0096-3445.137.1.73

von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-based and exemplar-based processes compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 867–889. doi:10.1037/a0015501

Weaver, E. A., & Stewart, T. R. (2012). Dimensions of judgment: Factor analysis of individual differences. *Journal of Behavioral Decision Making*, *25*, 402–413. doi:10.1002/bdm.748

Westbrook, J. I., Woods, A., Rob, M. I., Dunsmuir, W. T. M., & Day, R. O. (2010). Association of interruptions with an increased risk and severity of medication administration errors. *Archives of Internal Medicine*, *170*, 683–690. doi:10.1001/archinternmed.2010.65

Worthy, D. A., Markman, A. B., & Maddox, W. T. (2009). What is pressure? Evidence for social pressure as a type of regulatory focus. *Psychonomic Bulletin & Review, 16*, 344–349. doi:10.3758/PBR.16.2.344

Yogev-Seligmann, G., Hausdorff, J. M., & Giladi, N. (2008). The role of executive function and attention in gait. *Movement Disorders, 23*, 329–342. doi:10.1002/mds.21720

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition, 34*, 387–398. doi:10.3758/BF03193416

Zeithamova, D., & Maddox, W. T. (2007). The role of visuo-spatial and verbal working memory in perceptual category learning. *Memory & Cognition, 35*, 1380–1398. doi:10.3758/BF03193609

Supplemental Material

Supplemental Material A: Cognitive Modeling of Judgment Strategies

To understand the cognitive strategies underlying participants' judgments, we fitted a similarity-based judgment model (an exemplar model) and a rule-based judgment model (a linear model) to each participant individually. Performance of the exemplar model (with one sensitivity parameter, *h*) and the linear model (with six parameters, one for each cue and a constant intercept) was compared to a baseline model that calculated each participant's mean judgment.

Similarity-Based Model

Exemplar models assume that the similarity between the new object (the probe) and all stored exemplars is a major determinant of judgment. This similarity $S(p,j)$ between probe p and exemplar j is a decay function of the distance between two objects (Nosofsky & Zaki, 1998):

$$S(p,j) = e^{-d_{pj}},$$
(A1)

where d_{pj} is the distance between the two objects. The distance was determined by the number of matching object features (or cues) weighted by a sensitivity parameter h:

$$d_{pj} = h\left(\sum_{i=1}^{I} \left| x_{pi} - x_{ji} \right| \right),$$
(A2)

where x_{pi} and x_{ji} are the cue values of the probe p and the exemplar j on dimension i. The sensitivity parameter h reflects the participant's ability to discriminate between specific exemplars. The judgment $\hat{c}_{p,\text{Ex}}$ for the probe p is the average of the criterion values c_j for stored exemplars j, weighted by their similarities (Juslin, Olsson, & Olsson, 2003):

$$\hat{c}_{p,\text{Ex}} = \frac{\sum_{j=1}^{J} S(p,j) \cdot c_j}{\sum_{j=1}^{J} S(p,j)} \, . \tag{A3}$$

We also fitted an exemplar model with five attention parameters (one for each cue) to each participant's judgments. On average, this more complex version of the exemplar model performed worse in generalization. For the sake of clarity, we restrict our report to an exemplar model with one sensitivity parameter.

Rule-Based Model

The linear model assumes that the final criterion estimate $\hat{c}_{p,\text{Rule}}$ of an object p is a linear additive function of the object's cue values x_{pi}:

$$\hat{c}_{p,\text{Rule}} = k + \sum_{i=1}^{I} w_i \cdot x_{pi} \, , \tag{A4}$$

where w_i are the cue weights for each dimension i and k is a constant intercept.

Selection of Training and Validation Items

We constructed two different item sets for the training and the test phase: First, we generated 100 training sets with 16 randomly selected items and determined their criterion values according to

$$C = 0.2 \, e^{(22 x_1 + 20 x_2 + 17 x_3 + 15 x_4 + 12 x_5)/15} \tag{A5}$$

where C is the criterion ranging from 0 to 62 and x_1 to x_5 are the cue values. The remaining 16 items formed the validation set. From these training–validation set combinations we then selected all training sets that could not be solved by a rule-based strategy, that is, a linear model fitted these training sets worse than an exemplar model. Based on these training sets, we derived model predictions for the validation sets. We aimed for a validation set in which the exemplar model made more accurate predictions than the linear model, that is, in which a similarity-based strategy should lead to a higher judgment accuracy. Additionally, the final validation set was selected so that it strongly discriminated between the models' predictions.

Generalization Procedure

Since the models varied in the number of free parameters, that is, in model complexity, model performance was evaluated by a generalization test (Busemeyer & Wang, 2000): In a generalization test, the complete set of items is split into a calibration set and a validation set. The model parameters are then estimated from the calibration set and used to make new predictions for the validation set. The discrepancy between these new predictions and the actual data gives an index of model fit. A compelling advantage of this methodology is that it accounts for model complexity.

We fitted the models to each participant's judgments in the last three training blocks. Subsequently, the estimated parameters were used to predict each participant's mean judgment for each Sonic in the test phase. Model fit, the deviation between participants' mean judgments and the models' predictions, was measured in RMSD. After excluding participants best described by a baseline model, we determined a strategy weight w_s, indicating if the linear or the exemplar model described participants' judgments better:

$$\hat{c}_p = w_s \cdot \hat{c}_{p,\mathrm{Ex}} + (1 - w_s) \cdot \hat{c}_{p,\mathrm{Rule}}. \tag{A6}$$

Based on each model's optimal parameters, the strategy weight weighs the predictions of the exemplar model $\hat{c}_{p,\mathrm{Ex}}$ and the linear model $\hat{c}_{p,\mathrm{Rule}}$ by minimizing the deviation between the predicted judgments $\hat{c}_p$ and participants' judgments in the test phase. A weight larger than .5 indicates a higher probability of using the exemplar model; a weight smaller than .5 indicates a higher probability of using the linear model.

Results

Without cognitive load, both the linear model and the exemplar model described participants' judgments better than a baseline model, $t(29) = 2.89$, $d = 0.34$, $p = .007$ and $t(29) = 3.11$, $d = 0.34$, $p = .004$, respectively (see Table A1 for model fits and Table A2 for classification results). The exemplar model and the linear model could not be distinguished

from each other, $t(29) = 0.05$, $d < 0.01$, $p = .962$. Under low cognitive load, the exemplar

model reached a better fit than the linear model, $t(28) = 4.19$, $d = 0.70$, $p < .001$, and the

baseline model, $t(28) = 3.96$, $d = 0.70$, $p < .001$. The same pattern emerged under high

cognitive load: The exemplar model made more accurate predictions than the linear model,

$t(29) = 5.46$, $d = 0.90$, $p < .001$, and a baseline model, $t(29) = 3.67$, $d = 0.85$, $p = .001$.

Table A1

Cognitive Models' Goodness-of-Fit Measured in Mean Root Mean Square Deviations

(RMSDs) Separately for the Cognitive Load Conditions (Standard Deviations in Parentheses)

	Cognitive model		
Condition	Baseline model	Linear model	Exemplar model
No cognitive load	12.42 (5.74)	10.73 (3.97)	10.71 (3.99)
Low cognitive load	11.67 (3.65)	11.37 (2.85)	9.42 (2.74)
High cognitive load	10.89 (3.23)	11.67 (3.28)	8.87 (2.94)

Table A2

Classification of Participants According to the Cognitive Models (Percentages in

Parentheses) Separately for the Cognitive Load Conditions

	Cognitive model		
Condition	Baseline model	Linear model	Exemplar model
No cognitive load	4 (13%)	13 (43%)	13 (43%)
Low cognitive load	5 (17%)	5 (17%)	19 (66%)
High cognitive load	10 (33%)	2 (7%)	18 (60%)

Supplemental Material B: Extension to a Linear Judgment Task

In this study we investigated if cognitive load increased switching to similarity-based strategies in a linear task in which switching should hurt performance. To investigate this question we replicated the experiment with a commonly used linear judgment task (Mata, von Helversen, Karlsson, & Cüpper, 2012).

Method

Participants. Sixty participants (35 women, M_{age} = 25 years, SD_{age} = 7 years) took part in our study, 30 in each condition. Participants received 17 CHF per hour and a performance-contingent bonus (M = 5.4 CHF). One participant did not perform the operation span task due to a mistake of the experimenter.

Design. The task was to estimate a continuous criterion based on four binary cues. The criterion C was a linear additive function of the cue values according to

$$C = 4x_1 + 3x_2 + 2x_3 + x_4 + 10, \tag{B1}$$

where C is the criterion ranging from 10 to 20 and x_1 to x_4 are the cue values. Participants were randomly assigned to either a control condition or a high cognitive load condition. The control condition replicated Mata et al.'s (2012) judgment task. In the high cognitive load condition, however, participants solved a concurrent memory task.

Material. For the judgment task, we used the same cover story as in the main experiment. The Sonics' appearance varied on four binary cues: Hair, nose, ears, and body. Table B1 illustrates the task structure: The binary cues could take the cue values zero or one and varied in their cue weights. The cue weights were randomly assigned to the four pictorial cues, as were the cue values (zero or one) to the features (e.g., spiky vs. dreadlocks).

Procedure. Participants first completed an operation span task (Unsworth, Heitz, Schrock, & Engle, 2005). Afterward, participants solved the linear judgment task. As in the first study, the task consisted of a training and a test phase. During the training phase, participants repeatedly judged 10 training items. The training phase ended when a learning

criterion had been reached (after at least 8 and at most 16 blocks). Participants met this learning criterion when judgment accuracy fell below 1.5 RMSD. Participants in the high cognitive load condition additionally saw four consonants before the Sonic appeared on screen. Consonants were presented sequentially, each for 2 s. After participants received feedback about their judgment they were asked to recall the letters in their presentation order. In the test phase, participants judged 16 Sonics four times without feedback or cognitive load.

To increase their motivation, participants received a performance-contingent payment. In each trial, participants earned 10 points for a correct answer (corresponding to 0.05 CHF). If their judgments deviated by 1 from the correct answer, participants received 5 points. If their judgment deviated by more than 1 from the correct answer, participants did not receive any points. Participants under cognitive load received an additional point for a correct recall of the letters. Yet, when they could not recall the letters they did not earn any points for the memory *or* the judgment task. Additionally, participants were paid a bonus of 3 CHF if they reached the learning criterion for the judgment task within 16 training blocks.

Results

Adherence to cognitive load. As in the main experiment, letter recall was high. Under high cognitive load, participants recalled 91.0% (SD = 7.8%) of the letter sequences correctly. Higher judgment accuracy was related to better letter recall, $r(30)$ = -.37, p = .046, indicating no trade-off between the judgment and the memory task.

Working memory capacity. Working memory capacity in the high cognitive load condition (M = 38.1, SD = 18.6) was comparable to working memory capacity in the no load condition (M = 41.4, SD = 16.1), $t(57)$ = 0.73, p = .469.

Judgment performance. Overall, the majority of the participants (62%) mastered the task and reached the learning criterion (10 did not reach the criterion in the control condition and 13 in the high load condition). To learn the task, participants in the control condition needed as many blocks (M = 12.3, SD = 6.3) as participants in the high cognitive

load condition ($M = 12.3$, $SD = 3.6$). The training block data violated normality assumptions, thus, we conducted a Mann–Whitney U test to test for differences between the cognitive load conditions. This test revealed no differences between the cognitive load conditions, $U = 448$, $p = .975$. Also, judgment accuracy in the last training block did not differ between participants without cognitive load ($M = 1.67$, $SD = 1.12$) or under high cognitive load, $M = 1.84$, $SD = 1.04$, $t(58) = 0.60$, $p = .552$. In the test phase, judgments for trainings items were as accurate for participants without cognitive load ($M = 1.71$, $SD = 0.89$) as for participants under high cognitive load ($M = 1.75$, $SD = 0.92$), $t(58) = 0.13$, $p = .896$. Also, judgment accuracy for validation items did not differ between the cognitive load ($M = 2.32$, $SD = .97$) and the control condition ($M = 2.15$, $SD = 0.93$), $t(58) = 0.68$, $p = .497$. These results suggest that even under high cognitive load, participants learned the task well. In contrast to the nonlinear judgment task, however, cognitive load did not promote judgments for validation items.

Additional analysis including working memory capacity indicated that participants with a higher working memory capacity learned to solve the task more easily. In the high load condition, participants with a higher working memory capacity needed fewer training blocks to learn the task, Kendall's $\tau(28) = -.29$, $p = .041$. Without cognitive load, however, working memory capacity was not related to the number of training blocks, Kendall's $\tau(27) = -.12$, $p = .402$. Next, we included working memory capacity as a covariate when analyzing judgment performance. A higher working memory capacity marginally increased judgment accuracy in the last training block, $F(1,56) = 2.99$, $\eta^2 = .05$, $p = .089$. In the test phase, higher working memory capacity led to slightly better judgments for training items, $F(1,56) = 3.95$, $\eta^2 = .07$, $p = .052$, but not for validation items, $F(1,56) = 1.87$, $\eta^2 = .03$, $p = .178$. Thus, although a high working memory capacity slightly facilitated learning, it did not improve performance on validation items.

Cognitive modeling of judgment strategies. To test if people switched to a similarity-based strategy we followed the same approach as in the first study. We first fitted an exemplar model (similarity-based judgment strategy), a linear model (rule-based judgment strategy), and a baseline model to participants' individual judgments during the training phase (see Table B2 for model fits and Table B3 for classification results). In both conditions, the exemplar model and the linear model outperformed a baseline model in predicting participants' judgments. However, while in the control condition the linear model could predict participants' judgments marginally better than the exemplar model, $t(29) = 1.77$, $d = 0.33$, $p = .088$, there was no difference in model fits in the high cognitive load condition between the exemplar and the linear model. Further, we excluded participants following the baseline model. Then we estimated the strategy weight to capture how much participants relied on an exemplar versus a linear model. This strategy weight was slightly higher in the high load condition ($M = .53$; $SE = .07$) than in the control condition ($M = .33$, $SE = .07$), $t(49) = 1.91$, $d = 0.54$, $p = .061$. This result indicates that participants under high cognitive load were more likely to follow a similarity-based strategy.

Besides cognitive load, working memory capacity may influence strategy choice. Thus, we regressed strategy weight on working memory capacity, including cognitive load as a covariate. Working memory capacity did not predict strategy weight beyond cognitive load, $b = 0.001$, $SE = 0.003$, $t(47) = 0.33$, $p = .747$.

Judgment accuracy and cognitive models. In our first study, we showed that following a similarity-based strategy can benefit judgment performance in a nonlinear task. In a linear judgment task, however, reliance on an exemplar-based strategy should harm judgment performance. To test this assumption, we conducted a regression analysis on judgment performance for validation items with strategy weight as the independent variable. Indeed, a higher strategy weight predicted lower judgment performance on validation items, b

= -0.87, $SE = 0.32$, $t(49) = 2.73$, $p = .009$, $R^2 = .13$. This result suggests that using a similarity-based strategy harmed judgments in the linear task.

Table B1

Cue and Criterion Values of Training and Validation Items in the Linear Judgment Task

Cue 1	Cue 2	Cue 3	Cue 4	Criterion	Item set
0	0	0	0	10	Validation
0	0	0	1	11	Training
0	0	1	0	12	Training
0	0	1	1	13	Training
0	1	0	0	13	Training
0	1	0	1	14	Training
0	1	1	0	15	Validation
0	1	1	1	16	Training
1	0	0	0	14	Training
1	0	0	1	15	Validation
1	0	1	0	16	Training
1	0	1	1	17	Training
1	1	0	0	17	Validation
1	1	0	1	18	Validation
1	1	1	0	19	Validation
1	1	1	1	20	Training

Note: Training items were presented in the training and the test phase. Validation items only appeared during the test phase.

Table B2

Cognitive Models' Goodness-of-Fit Measured in Mean Root Mean Square Deviations

(RMSDs) Separately for the Cognitive Load Conditions (Standard Deviations in Parentheses)

	Cognitive model		
Condition	Baseline model	Linear model	Exemplar model
No cognitive load	2.00 (0.55)	1.29 (0.57)	1.45 (0.37)
High cognitive load	2.10 (0.41)	1.61 (0.71)	1.55 (0.51)

Table B3

Classification of Participants According to the Cognitive Models (Percentages in

Parentheses) Separately for the Cognitive Load Conditions

	Cognitive model		
Condition	Baseline model	Linear model	Exemplar model
No cognitive load	5 (17%)	16 (53%)	9 (30%)
High cognitive load	4 (13%)	12 (40%)	14 (46%)

References

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*, 171-189. doi: 10.1006/jmps.1999.1282

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132*, 133-156. doi: 10.1037/0096-3445.132.1.133

Mata, R., Helversen, B. von, Karlsson, L., & Cüpper, L. (2012). Adult age differences in categorization and multiple-cue judgment. *Developmental Psychology, 48*, 1188-1201. doi: 10.1037/a0026084

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247-255. doi: 10.1111/1467-9280.00051

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498-505. doi:10.3758/BF03192720

From rules to exemplars: Similar task features shape judgment and categorization processes

Janina A. Hoffmann, Bettina von Helversen, and Jörg Rieskamp

University of Basel, Department of Psychology, Basel, Switzerland

Abstract

The distinction between similarity-based and rule-based strategies has instigated a large body of research in categorization and judgment. Although the conditions guiding processing shifts are increasingly well documented within both tasks, it is largely unclear how factors which influence strategy choice in one task transfer to the other task. In two studies, we aimed to integrate research from these two fields by investigating how task components affecting strategy choices in judgment or categorization influence strategy choice across tasks. Specifically, we investigated how the number of dimensions, the functional relation between cue and criterions, and individual preferences influence strategy choice in categorization and judgment. In two experiments we varied the type of task (categorization or judgment) within participants and task structure between participants, comparing a one-dimensional linear to a multidimensional linear and two multidimensional nonlinear tasks. In both categorization and judgment more participants relied on similarity-based strategies when more dimensions had to be integrated and when the functional relationship was nonlinear suggesting that strategic shifts may be driven by task complexity. With more complex tasks people more frequently switched strategies between tasks suggesting that individual preferences may be overruled by task characteristics.

Keywords: Judgment; categorization; cognitive processes

On many occasions in everyday life, the same task can demand a coarse classification or a more fine-grained judgment. When applying for a job, for instance, the applicant may sort the jobs into broad categories such as "highly interesting" or "not interesting at all". Alternatively, the applicant may judge on a more fine-grained scale how attractive the jobs are. Prototypical tasks to investigate judgments and categorizations share indeed some commonalities (Juslin, Olsson, & Olsson, 2003). Beyond sharing task characteristics, both research fields identified two main types of strategies people use to judge or classify objects (Erickson & Kruschke, 1998; Juslin, Olsson et al., 2003; McDaniel, Cahill, Robbins, & Wiener, 2013; von Helversen & Rieskamp, 2008, 2009): similarity-based strategies and rule-based strategies. These strategies make different assumptions about the way knowledge is represented and the cognitive processes underlying judgments and categorizations (Hahn & Chater, 1998; Juslin, Olsson et al., 2003). In general, it is assumed that similarity-based strategies base inferences upon on a comparison with concrete instances stored in memory, whereas rule-based strategies rely upon explicit abstraction of knowledge (Hahn & Chater, 1998). Although giving a coarse or a fine-grained response should not influence the cognitive processes that underlie the response, the two different research traditions have mostly described categorizations by similarity-based strategies, whereas judgment processes have been characterized as rule-based (Juslin, Olsson et al., 2003; von Helversen & Rieskamp, 2009). Confirming this characterization, past research suggests that people frequently shift from rule-based strategies in judgment to similarity-based strategies in categorization (Juslin, Olsson et al., 2003; Pachur & Olsson, 2012; von Helversen, Karlsson, Mata, & Wilke, 2013; von Helversen, Mata, & Olsson, 2010). However, it is unclear if this strategy shift generalizes across various task characteristics.

In fact, people can make use of both rule-based and similarity-based strategies in judgment (Hoffmann, von Helversen, & Rieskamp, 2013a; Juslin, Karlsson, & Olsson, 2008; Karlsson, Juslin, & Olsson, 2007) and categorization (Juslin, Jones, Olsson, & Winman,

2003; Rouder & Ratcliff, 2004) with people choosing a strategy based upon the strategies' accuracy and the effort associated with executing these strategies (Beach & Mitchell, 1978; Payne, Bettman, & Johnson, 1993; Rieskamp & Otto, 2006). Across judgment and categorization, however, different task characteristics have been identified as determinants of strategy choice. Whereas categorization research has focused on the number of cues that need to be integrated (e.g. Ashby, Maddox, & Bohil, 2002; Filoteo, Lauritzen, & Maddox, 2010; Maddox & Ashby, 2004; Zeithamova & Maddox, 2006), studies in judgment have focused on the functional relationship between the cues and the criterion as the most important factor influencing strategy choice (Hoffmann et al., 2013a; Karlsson et al., 2007; Juslin et al., 2008).

In the current paper we aim to integrate judgment and categorization research by investigating how the number of cues (or dimensions) and the functional relation between cue and criterion affect strategy choice across categorization and judgment tasks. Specifically, we suggest that people switch strategies in both tasks in response to task characteristics and these shifts can be explained by the relative accuracy of the strategies and the effort necessary to execute the strategies successfully. In the following we will first review past research on how the number of cues and the functional relation affect reliance on rule-based and similarity-based strategies in judgment and categorization and, second, outline how a framework based on strategy accuracy and effort can explain these strategy choices. Finally, we will report two experimental studies to test our hypotheses.

Rule-based and Similarity-based Strategies in Categorization and Judgment

In general rule-based strategies are assumed to involve controlled processes that rely on abstracted knowledge (Juslin et al., 2008; Karlsson, Juslin, & Olsson, 2008). A typical rule-based strategy, the cue abstraction strategy (Juslin, Olsson et al., 2003; Juslin, Jones et al. 2003), assumes that people abstract how each cue relates to the criterion, that is they try to find out the importance of each cue. The judgment is the sum of the cue values, weighted by their importance. A job applicant may, for instance, try to figure out how much he appreciates

challenging tasks or a high wage and assign a high weight to wage. Accordingly, the job applicant will rate jobs as more attractive the higher the jobs are paid. In a similar way, a person may follow the rule that all jobs are classified as interesting that do pay a minimum salary. Hence the probability of classifying the job as interesting should increase with increasing wage. Rule-based strategies proposed in the literature vary in their complexity from rules considering only one cue or two cues (Ashby & Maddox, 2005; Erickson & Kruschke, 1998; Nosofsky, Little, & Denton, 2011; Nosofsky, Palmeri, & McKinley, 1994) to linear rules with several cues (Juslin, Olsson et al., 2003; Juslin et al., 2008; Newell, Weston, Tunney, & Shanks, 2009; Platzer & Bröder, 2013). Are there limits in the complexity rules can take? By definition, rule-based strategies rely on abstract explicit knowledge implying that rules can be verbalized (Ashby & Maddox, 2005). Evidence suggests that people build up task knowledge in linear tasks with multiple cues indicating that people can follow linear rules and possess insight in the rule-based process (Lagnado, Newell, Kahan, & Shanks, 2006). More complex nonlinear rules, however, have often been rejected based on theoretical and empirical grounds (Busemeyer, Byun, DeLosh, & McDaniel, 1997; Juslin et al., 2008).

In contrast, a typical similarity-based strategy, the exemplar model, assumes that the similarity to past instances is used to make a categorization or judgment (Juslin, Olsson et al., 2003; Medin & Schaffer, 1978; Nosofsky, 1988). These exemplar models assume that all previously encountered objects (the exemplars) are stored in long-term memory along with their associated categories. When categorizing a new object (the probe), past exemplars are retrieved from memory and the probe is compared to all exemplars stored in memory. The more similar the probe is to a past exemplar, the more likely the probe will be classified as belonging to the same category. For instance, when categorizing a new job offer, people may remind themselves of all jobs they had in the past. If the job applicant liked jobs with customer interaction in the past, the applicant will probably also judge a new job offer as attractive that requires customer contact.

Factors Encouraging Shifts between Rule-based and Similarity-based Strategies

Number of Cues

The categorization literature has suggested that people approach a categorization task by testing simple rules that consider only one or two dimensions. In case these rules are not successful, people switch to similarity-based strategies (Erickson & Kruschke, 1998; Nosofsky et al., 1994). For instance, Nosofsky et al. (1994) suggested that people test simple one- or two-dimensional rules when learning categorization tasks, but store exceptions in memory if the rules do not work. Similarly, Erickson and Kruschke (1998) suggested that people simultaneously process rules and exemplars, but restricted the rules tested to one dimension. Furthermore, people seem to process categorization tasks differently that can be solved by a simple one- or two-dimensional rule compared to categorization tasks that require information integration[1] (Ashby et al., 2002; Filoteo et al., 2010; Maddox & Ashby, 2004; Zeithamova & Maddox, 2006). In sum, this suggests that the number of cues is an important factor driving rule-based or similarity-based processing.

Likewise, meta-analyses in judgment identified the number of cues as one major factor determining judgment performance (Karelaia & Hogarth, 2008; Kaufmann & Athanasou, 2009). If more cues have to be considered for making a judgment, judgment performance decreases (Karelaia & Hogarth, 2008). Kareleia and Hogarth (2008) reasoned that this decrease can be explained by a decreasing match between the linear cue combination of the judge and the linear model of the environment. Instead people may follow more complex compound cue strategies if the number of cues increases (Einhorn, 1971). Hence, it is possible that an increasing number of cues foster similarity-based strategies. As a factor influencing strategy choice, however, the number of cues has been — to our knowledge — mostly neglected.

Functional Relationship between Cues and Criterion

The main factor influencing strategy shifts in judgment is the functional relation between the cues and the criterion (Hoffmann et al., 2013a; Juslin et al., 2008; Karlsson et al., 2007; von Helversen & Rieskamp, 2008). Indeed, the majority of research suggests that people rely more on similarity-based strategies if the task cannot be solved by a linear rule, for instance, if the criterion is a multiplicative function of the cues (Hoffmann et al., 2013a; Juslin et al., 2008). One exception to this pattern constitute quadratic task structures in which the criterion is a quadratic function of the cues. Because the same criterion value is associated with multiple, but dissimilar exemplars, neither similarity-based strategies nor rule-based strategies yield to good performance early in training and people drop back to the default, but useless cue abstraction process (Karlsson et al., 2008; Olsson, Enkvist, & Juslin, 2006).

The functional relationship between cue and criterion has also been studied in function learning tasks in which people learn to predict a continuous criterion based on one cue with varying functional relationships between cue and criterion. Overall, past research suggests that linear relationships are learnt faster than exponential or quadratic functions (Busemeyer et al., 1997; DeLosh, Busemeyer, & McDaniel, 1997). In addition, rule-based function learning models fare well at predicting extrapolation for linear functions, but fail on exponential or quadratic functions (De Losh et al., 1997; McDaniel & Busemeyer, 2005). However, extrapolation frequently follows an approximately linear function and associative, similarity-based models only account successfully for extrapolation in more complex tasks if they incorporate a linear, rule-based extrapolation mechanism (De Losh et al., 1997; McDaniel & Busemeyer, 2005).

In categorization, nonlinear or quadratic relationships are also learned more slowly and less accurately than linear relationships (Ashby & Gott, 1988), but people can reach near-optimal performance when learning nonlinear bounds (Ashby & Maddox, 1992). Overall, however functional relations have been rarely considered as a factor influencing rule-or

similarity-based processing or these comparisons have led to inconclusive results (Maddox & Ashby, 1993; McKinley & Nosofsky, 1995, 1996; Wills & Pothos, 2011).

In sum, past research suggests that the number of cues is an important factor influencing strategy shifts in categorization but research in judgment is scarce. Similarly, the functional relationship has been identified as an important factor influencing strategy selection in judgment, but only rarely considered in categorization. A framework to understand strategy selection has been offered by contingency approaches (Beach & Mitchell, 1978; Payne et al., 1993). These approaches assume that people select among a set of decision strategies by trading off the accuracy that can be achieved by following a strategy against the effort that needs to be invested in learning and executing a strategy. The functional relationship between cues and criterion may limit a strategy's accuracy, whereas a higher number of cues may increase the effort necessary to execute a strategy.

Understanding Strategy Choice As an Accuracy-Effort Trade-Off

A large body of research suggests that strategy choices are driven by the accuracy of the strategies. Task feedback reinforces the better performing strategy and thus allows the decision maker to adapt their strategies to the demands of the task (Juslin, Olsson et al., 2003; Rieskamp & Otto, 2006). Likewise, the relative accuracy of strategies has been suggested as the main mechanism underlying strategy choices in judgment and categorization. For instance, models assuming that people switch between rule-based and similarity-based strategies frequently assume that the probability of a given process depends on its accuracy (Erickson & Kruschke, 1998; Juslin et al., 2008; Nosofsky et al., 1994). Indeed, people often prefer the more accurate strategy (Filoteo et al., 2010; Lewandowsky, Yang, Newell, & Kalish, 2012). Moreover, even if the task structure suddenly changes, people are able to adapt — to some degree —decision strategies to the decision task based upon task feedback (Kämmer, Gaissmaier, & Czienskowski, 2013; Rieskamp & Otto, 2006). Similarly, people are more frequently relying on an exemplar-based strategy in multiplicative judgment tasks, because

the cue abstraction model does not achieve accurate judgments (Hoffmann et al., 2013a; Juslin et al., 2008; von Helversen et al., 2013).

Besides accuracy also the effort with which the strategies can be learnt has been identified as an important factor in strategy choice (Beach & Mitchell, 1978; Payne et al., 1993; Rieskamp & Otto, 2006). For instance, if people can easily discriminate past exemplars, it is easier to store these exemplars and, accordingly, people tend to rely more strongly upon similarity-based strategies (Rouder & Ratcliff, 2004). In contrast, if cue directions are known, less effort needs to be invested in abstracting cue weights so that more people rely on rule-based learning (Platzer & Bröder, 2012; von Helversen et al., 2013; Newell et al., 2009). Furthermore, also different forms of feedback may increase or reduce the effort associated with abstracting rules. Pachur and Olsson (2012), for instance, found that learning which of two objects has a higher criterion value enhances reliance on cue abstraction processes, possibly because people focus on how differences in cue values are associated with differences in judgment criteria —an important step in cue abstraction. Binary feedback in categorization, however, makes abstracting the correct rule more difficult, resulting in more people switching from a cue abstraction strategy in judgment to similarity-based categorization strategies (Juslin, Olsson et al., 2003; Karlsson et al., 2008; von Helversen et al., 2010).

Lastly, people may also build up initial preferences for specific strategies over time because they learn to associate each strategy with its achieved accuracy and the involved effort (Beach & Mitchell, 1978; Rieskamp & Otto, 2006). It has been suggested recently that these individual preferences for rule- or exemplar-based strategies may be rather stable and transfer across tasks (McDaniel et al., 2013). In particular, McDaniel et al. (2013) found that people identified as rule-learners in complex function learning tasks generalize their performance more successfully than exemplar-learners to new items in abstract categorization tasks. Furthermore, these preferences may be linked to individual differences such as memory

ability (Hoffmann, von Helversen, & Rieskamp, 2013b; McDaniel et al., 2013) or age (Mata, von Helversen, Karlsson, & Cüpper, 2012).

Hypotheses

Taken together, the relative accuracy and effort of the strategies can be used to predict how the number of dimensions and the functional relationship between the cues and the criterion should influence participants' strategies in judgment and categorization. Specifically, in judgment, one would expect that most participants should rely on a rule-based strategy in a one-dimensional linear (OLIN) task, because the rule-based strategy is accurate and easy to learn. In a multidimensional linear (MLIN) task a cue abstraction strategy is correct, but more cues have to be considered than in the OLIN task making the rule-based strategy more difficult to learn. Accordingly, some people may rely upon initial preferences and default to an exemplar-based strategy. Finally, in a multidimensional multiplicative (MMULT) task, the majority should switch to an exemplar strategy because now a cue abstraction strategy fails.

In categorization a similar pattern is expected: Specifically, one would expect the largest number of rule-users in an OLIN task. In MLIN tasks, however, more cues hinder correctly abstracting the weights for more complex rule-based strategies. Importantly, binary feedback in categorization — in comparison to more fine-grained feedback in judgment — further complicates cue abstraction leading to an even higher percentage of exemplar users in the MLIN categorization task than in judgment. Finally, in a MMULT task, this shift to exemplar-based strategies should be even more pronounced because any linear cue abstraction strategy fails to solve this task.

Lastly, if people have stable preferences for exemplar-based or rule-based strategies people should tend to rely on the same strategy in both tasks. Specifically, if one used an exemplar-based strategy in the first task, the conditional probability of using an exemplar-based strategy in the second task should be close to 1. Likewise, if one used a rule-based

strategy in the first task, the conditional probability of using a rule-based strategy in the second task should be close to 1.

Study 1

To test the influence of these factors, we conducted an experiment in which 96 participants solved both a categorization and a multiple-cue judgment task with the same task structure allowing us to investigate preferences across tasks. In addition, we varied the task structure on three levels, comparing an OLIN to a MLIN and a MMULT task to investigate if the results found in categorization and judgment can be generalized across tasks.

Method

Participants.

Ninety-six participants (76 females, $M_{Age} = 23.7$, $SD_{Age} = 5.9$) were recruited from the University of Basel. Participants received course credit or a book certificate (worth 25 Swiss Francs, CHF) for participating in the experiment. In addition, they could earn a bonus of 3 CHF in each task and had the opportunity to win one of six Amazon vouchers (worth 25 CHF each).

Design and material.

We used two different cover stories for the categorization and the multiple-cue judgment task. One cover story asked participants to judge the toxicity of a bug: In the multiple-cue judgment task participants estimated how toxic a bug was on a scale from 0 to 50, whereas in the categorization task participants classified the bug as toxic or harmless. The other cover story asked participants to judge how successful comic figures were at catching small animals: In the multiple-cue judgment task, participants judged how many small animals the comic figure caught on a scale from 0 to 50, whereas they classified the comic figure as catching few or many animals in the categorization task.

The stimuli for the two cover stories consisted of pictures of either bugs or comic figures. These bugs and comic figures varied on four different continuous cues. The bugs

varied on the length of their legs, their antennae and their wings, and the number of points on their back. The comic figures had different sizes of their ears and their nose, a different number of hairs and stripes on their shirt. These pictorial cues could be used to predict the criterion (the toxicity of a bug or the success of the comic figure).

To manipulate task complexity, we varied how these cues had to be combined to form the judgment criterion. In the MLIN task, the criterion y_{MLIN} was a linear, additive function of the cues:

$$y_{MLIN} = 4\,c_1 + 3\,c_2 + 2\,c_3 + c_4, \tag{1}$$

where c_1 to c_4 are the cue values ranging from 0 to 5. According to their cue weights, c_1 reflects the most important cue and c_4 the least important one.

In the OLIN task only one cue predicted the judgment criterion y_{OLIN}:

$$y_{OLIN} = 10\,c_3. \tag{2}$$

Finally, in the MMULT task, the function generating the criterion y_{MMULT} included a multiplicative combination of the cues:

$$y_{MMULT} = \frac{4c_1 + 3c_2 + 2c_3 + c_4 + 2c_1c_2c_3 + c_2c_3c_4}{8.5} \tag{3}$$

In the categorization tasks, the criterion was not continuous anymore, but binary. This binary criterion was created by a median split on the corresponding judgment criterion for all possible items. Sonics (or bugs) with criterion values above the median were classified as catching many animals (or as toxic). Sonics (or bugs) with criterion values below the median were classified as catching few animals (or as harmless). This median split creates a linear category boundary in the OLIN and the MLIN task and a nonlinear category boundary in the MMULT task.

In all tasks, the cues were randomly assigned to the pictorial cues (e.g., ears or nose). Higher cue values, however, were always associated with more salient pictorial cue features.

For instance, a cue value of zero corresponded to a bug without points on the back and a cue value of five to a bug with five points on its back.

From all possible items, we constructed a training set and a validation set. First, we generated 1000 trainings including 25 training items. Second, we selected one training set fulfilling two criteria: (a) One or two dimensional rules should not lead to a high accuracy in the multidimensional categorization tasks. (b) Rule-based processes should solve the MMULT judgment and categorization task worse than exemplar-based processes, that is, a (log-) linear regression fitted the training set worse than an exemplar model. Next, we generated 100 validation sets consisting of 15 training items. Finally, we selected a training-validation set combination in which the validation set strongly discriminated between the models' predictions in all judgment and categorization tasks. Table 1 depicts the final training set and Table 2 the validation set.

Procedure.

Participants solved both a categorization and a judgment task with the same task structure. Participants were randomly assigned to three different task structures: OLIN, LIN, or MMULT tasks. Thirty-two participants were assigned to each condition. The assignment of the cover stories to the tasks and the order of the tasks were counterbalanced within each condition.

Both tasks consisted of a training phase and a test phase. During the training phase, participants learned to predict the criterion value (or the category) for 1 of 25 training items. In each trial they first estimated the criterion or categorized the item. Afterwards they received feedback about their own answer, the correct outcome and the points they earned. In a training block, all 25 training items were presented in random order. After 10 training blocks the training phase ended and participants moved on to the test phase. In this test phase, participants judged all 15 new validation items four times without getting any feedback.

Participants were incentivized to achieve a high task performance. In each trial of the categorization tasks, participants could earn 20 points for a correct answer, 10 points for items that were classified with a probability of .5 to both categories, and 0 points for an incorrect answer. In the judgment tasks, participants earned more points the less their judgment j deviated from the correct criterion y:

$$\text{Points} = 20 - \frac{(j - y)^2}{7.625} \qquad (4)$$

This function was truncated so that participants could win at most 20 points and could not loose any points in each trial. The more points participants earned in a task the higher were their chances of winning an Amazon coupon for that task. In addition, participants could earn a bonus of 3 CHF in both tasks, if they reached 80% of the points in the last training block. In the categorization task, this learning criterion corresponded to 80% correct classifications. In the judgment task, judgment accuracy was measured in root-mean-square deviations (RMSD) between participants' judgments and the criterion. Participants reached the learning criterion if judgment accuracy was below 5.5 RMSD in the last training block.

Results

Performance in the categorization task.

Overall, participants solved the OLIN task more successfully than the MLIN or the MMULT task. Table 3 reports the mean percentage of errors in the last training block and the four test blocks. Participants made fewer errors in the OLIN task than in the MLIN or MMULT task in the last training block as well as in the test phase. Because the error rates in the OLIN task deviated from normality and variances were not homogeneous, we used non-parametric tests to test for the effect of the conditions. The number of cues affected how well people had learned the categorization task in the last training block, but not the functional relationship. Participants made fewer errors in the OLIN task than in the MLIN task (Mann-Whitney $U = 64.5$, $p < .001$) and in the MMULT task ($U = 85.5$, $p < .001$), but there was no

difference in the error rates between the MLIN and the MMULT task ($U = 499.5$, $p = .865$).[2]

Similarly, participants made fewer errors on the validation items in the OLIN task than in the

MLIN task ($U = 47$, $p < .001$) and in the MMULT task ($U = 60.5$, $p < .001$). But like in the

last training block, error rates did not differ between the MLIN and MMULT task ($U = 441.5$,

$p = .343$).

Performance in the judgment task

In the judgment task we measured accuracy as the RMSD between the criterion value

and participants' judgment. Similar to the categorization task, participants made— on average

— more accurate judgments in the OLIN task than in the MLIN or the MMULT task (see

Table 3) Again, judgment accuracy in the OLIN task was not normally distributed and

variance homogeneity was not given. Therefore, we relied on nonparametric tests to test the

differences between conditions. As in the categorization task, participants made more

accurate judgments in the OLIN task than in the MLIN task ($U = 200$, $p < .001$) or in the

MMULT task ($U = 222$, $p < .001$), but judgment accuracy did not differ between the MLIN

and the MMULT task ($U = 391$, $p = .104$).[3] Similarly, participants judged the validation items

more accurately in the OLIN task than in the MLIN task ($U = 131$, $p <.001$) or in the

MMULT task ($U = 159$, $p < .001$). Judgment accuracy did not differ between the MLIN and

the MMULT task ($U = 453$, $p = .428$).

Modeling of cognitive processes

To identify the cognitive processes people rely on in judgment and categorization, we

used a computational modeling approach. We compared how well three cognitive models

described participants' responses at the end of training and predicted participants' responses

in the test phase: a baseline model (estimating participants' mean judgment or the category

bias), a cue abstraction model modeled by a (log-) linear regression for rule-based strategies,

and an exemplar model with a free sensitivity parameter for similarity-based strategies.[4]

Cue abstraction model. We relied on a linear cue abstraction model as a prototypical rule-based strategy. The cue abstraction models can represent simple rule-based strategies relying on a single cue, but also allows more complex rules combining several cues in a linear additive fashion. It does not include nonlinear rules or interactions because there is little evidence that these can be learnt via a rule-based strategies (Busemeyer et al., 1997; Juslin et al., 2008)[5]. The cue abstraction processes can be mathematically described with linear regression models. Accordingly, the estimated criterion value $\hat{c}_p$ of an object p is the weighted sum of the cue values x_{pi},

$$\hat{c}_p = k + \sum_{i=1}^{I} w_i \cdot x_{pi}$$

(5)

where w_i are the cue weights for each cue i and k is a constant intercept.

The probability to classify an object to category b, $p(\hat{b} = 1)$, can be predicted by logistic regression models.

$$p(\hat{b} = 1) = \frac{e^{k + \sum_{i=1}^{I} w_i \cdot x_{pi}}}{1 + e^{k + \sum_{i=1}^{I} w_i \cdot x_{pi}}}.$$

(6)

The smoother logistic function accounts for random error in the decision making process (Juslin, Jones et al., 2003).

Exemplar model. In exemplar models, the similarity $S(p,j)$ between the probe p and exemplar j is an exponential decay function of the distances d_{pj} between the objects (Nosofsky & Zaki, 1998).

$$S(p,j) = e^{-d_{pj}}.$$

(7)

Thus, smaller distances between the probe p and exemplar j indicate a higher similarity between theses objects. To determine this distance, the cue values x_{pi} of probe p are compared to the cue values x_{ji} of exemplar j on all cues i. The more the cue values match each other, the smaller is the distance between the objects (Nosofsky & Johansen, 2000).

$$d_{pj} = h\left(\sum_{i=1}^{I}\left|x_{pi} - x_{ji}\right|\right).$$

(8)

The sensitivity parameter h determines how strongly similarity decays with distance. Smaller sensitivity parameters indicate a lower decline of similarity with distance.

The probability of categorization the probe p into response category b, $p(\hat{b} = 1)$, can then be determined calculating the similarity of probe p to all exemplars in category b and comparing it to the similarity of probe p to all exemplars (Nosofsky, 1988).

$$p(\hat{b} = 1) = \frac{\sum_{j=1}^{J} \beta \cdot S(p, j_{b=1})}{\sum_{j=1}^{J} \beta \cdot S(p, j_{b=1}) + \sum_{j=1}^{J} (1 - \beta) \cdot S(p, j_{b=0})}$$

(9)

The category bias β finally models how much people tend to respond with category b.

To account for judgment processes, Juslin, Olsson et al. (2003) assumed that the criterion value c_j of an exemplar is stored together with its cue values in memory. To estimate the criterion value of a new probe $\hat{c}_p$, the criterion values c_j for each exemplar are weighted by the similarities.

$$\hat{c}_p = \frac{\sum_{j=1}^{J} S(p, j) \cdot c_j}{\sum_{j=1}^{J} S(p, j)}$$

(10)

Model estimation and comparison. All models were fitted to participants' responses in the last three training blocks by minimizing the deviance -2LL, the negative summed log-likelihood L of the model given the data.

$$-2LL = -2 \cdot \sum \ln(L)$$

(11)

In the categorization task, the likelihood was the models' predicted probability of the chosen category. In the judgment task, we calculated the likelihood as the probability density of

participants' judgments j assuming a truncated normal distribution with the models' predicted

responses $\hat{c}_p$ as the mean of the normal distribution and a fitted standard deviation σ.[6]

$$L = \frac{\frac{1}{\sigma}\phi\left(j\middle|\hat{c}_p,\sigma\right)}{\Phi\left(50\middle|\hat{c}_p,\sigma\right) - \Phi\left(0\middle|\hat{c}_p,\sigma\right)} \tag{12}$$

This truncated normal distribution was chosen because it matched the response scale from 0

to 50.

To compare which model described participants' judgments better at the end of

training, we calculated the Bayesian Information Criterion (BIC; Schwarz, 1978). This model

selection criterion can be used to compare non-nested models. In addition, the BIC penalizes

overly complex models by accounting for the number of free model parameters k:

BIC = -2LL + k ln(n), (13)

where n denotes the number of observations. Smaller BIC values indicate a better model fit.

The estimated parameter values were then used to predict participants' average

responses on the validation items during the test phase. To determine model fit, we then

calculated the deviances based upon the difference between model predictions' and

participants responses. This generalization test corrects not only for model complexity in

terms of the number of free parameters, but it also corrects for functional complexity

(Busemeyer & Wang, 2000). Finally, we used this generalization test to classify participants

as following a cue abstraction model, an exemplar model, or a baseline model (Hoffmann et

al., 2013a).

Model fits and deviances.

To compare model fits in training and test, we relied upon Wilcoxon tests for paired

data, because BICs and deviances were not normally distributed. Categorizations at the end of

training were overall not well described by a baseline model (see Table 4 for BICs, deviances,

and strategy classification). In the OLIN task, the exemplar model accounted better for

participants' categorizations than the cue abstraction model ($z = 4.66$, $p < .001$), but did not distinguish between the cue abstraction and the exemplar model in the MLIN ($z = 1.42$, $p = .155$) or MMULT task ($z = -0.99$, $p = .322$). Overall, the baseline model could also not predict categorizations better than the cue abstraction or the exemplar model for validation items in the test phase. In the OLIN task, the cue abstraction model fared better at predicting categorizations than the exemplar model ($z = -4.26$, $p < .001$) suggesting that the BIC punished the cue abstraction model too harshly. In the MLIN task, however, the cue abstraction model could neither be distinguished from the baseline model ($z = 1.31$, $p = .191$) nor the exemplar model ($z = 0.37$, $p = .708$). Likewise, in the MMULT task, model deviances again did not disentangle the cue abstraction and the exemplar model ($z = 0.15$, $p = .881$). Accordingly, comparing average model fit did not suggest that rule- or exemplar-based processes dominated categorization behavior in the MLIN or MMULT task.

Like in the categorization task, a baseline model could not account for participants' judgments at the end of training (see Table 5 for BICs, deviances and strategy classification) The exemplar model described judgments more accurately than the cue abstraction model in the OLIN task ($z = 2.9$, $p = .003$) and the MMULT task ($z = 2.7$, $p = .007$). In the OLIN task, however, the two models could not be distinguished by BIC values ($z = 1.2$, $p = .239$). Mirroring the results from the training phase, the baseline model was also not able to predict participants' judgments in the test phase. In the OLIN task, the cue abstraction model also made more accurate predictions than the exemplar model ($z = 4.9$, $p < .001$). However, the generalization test could not discriminate between the cue abstraction model and the exemplar model in the MLIN task ($z = 1.1$, $p = .278$) or the MMULT task ($z = 1.8$, $p = .079$).

Strategy classification.

To investigate how the number of cues and the functional relationship influenced judgment and categorization strategies, we first classified each participant based upon the model deviances as best described by a cue abstraction, an exemplar, or a baseline model.

Descriptively, in the categorization tasks, most participants relied upon a rule-based strategy in the OLIN task (see Table 4 and Figure 1). Likewise, most participants followed a rule-based strategy in the MLIN task. In the MMULT task, however, one half of the participants were best described by an exemplar-based model, while the other half was best described by a cue abstraction model. The classification yielded a similar picture for the judgment task (see Table 5 and Figure 1): In the OLIN task, almost all participants were assigned to the cue abstraction model. The cue abstraction model still described most participants best in the MLIN task, whereas more than half of the participants were best described by the exemplar model in the MMULT task.

Next, we conducted a multivariate ordinal logistic regression analysis on the strategy classification in categorization and judgment. The independent variable type of task (categorization vs. judgment) was repeated within participants. In addition we included one variable coding the number of cues and a second variable coding the functional relationship. Overall, participants shifted more to exemplar-based processes from the OLIN to the MLIN task, $b = -2.28$, $SE = .39$, Wald $\chi^2(1) = 35.08$, $p < .001$, and again more from the MLIN to the MMULT task, $b = -0.92$, $SE = .39$, Wald $\chi^2(1) = 5.55$, $p = .019$, indicating that both the number of cues and the functional relationship led to a shift in cognitive processes. The type of task, however, did not affect the cognitive process, $b = -0.06$, $SE = .30$, Wald $\chi^2(1) = 0.04$, $p = .839$. Repeating the analysis only for the MLIN task neither indicated that participants shifted more to exemplar-based processes in the categorization task compared to the judgment task, $b = -0.56$, $SE = .49$, Wald $\chi^2(1) = 1.29$, $p = .256$. In sum, these results suggest that participants indeed adapted the cognitive process to the number of cues and the functional relationship between the cues and the criterion, but the type of task did not affect the process people relied.

Matching processes in judgment and categorization.

To find out how individual preferences for rule-based or exemplar-based learning affected shifts between cognitive processes, we classified participants in a first step as following a cue abstraction model in both tasks, as following an exemplar model in both tasks, or as shifting between strategies in both tasks irrespective of the type of task (judgment or categorization). Overall, the number of cues and the functional relationship changed shifting behavior significantly, $\chi^2(4)=10.06$, $p = .039$. While in the OLIN task most participants ($n = 26$) relied upon a cue abstraction model in both tasks, the number of participants following a cue abstraction model decreased in the MLIN task ($n = 15$) and the MMULT task ($n = 7$). By contrast, the number of participants assigned to the exemplar model in both tasks increased from the OLIN ($n = 0$), to the MLIN ($n = 6$), to the MMULT task ($n = 10$). However, also the number of participants shifting between processes increased from the OLIN ($n = 6$), to the MLIN ($n = 11$), to the MMULT task ($n = 15$).

Figure 2 depicts the conditional probability of following a cue abstraction model (an exemplar model) in the second task given that participants were best described by a cue abstraction model (an exemplar model) in the first task. In the OLIN task, participants were likely to stay with a cue abstraction model (CAM) in the second task if they were best described by a cue abstraction model in the first task, $p(\text{CAM}_{\text{Second}}| \text{CAM}_{\text{First}})$. In addition, they were unlikely to follow an exemplar model in the second task, even if they were best described by an exemplar model in the first task, $p(\text{Exemplar}_{\text{Second}}| \text{Exemplar}_{\text{First}})$. While $p(\text{CAM}_{\text{Second}}| \text{CAM}_{\text{First}})$ decreased in the MLIN task and even more in the MMULT task, $p(\text{Exemplar}_{\text{Second}}| \text{Exemplar}_{\text{First}})$ consistently increased from the OLIN to the MLIN task and even more in the MMULT task. However, the probabilities in the MMULT task are less distinct from each other and closer to .5 (a probability of .5 would be expected, if half of the participants shifted from a cue abstraction model in the first task to a different strategy in the second task), indicating that more cues and a more complex functional relationship make it more difficult to predict from the first task the cognitive processes underlying the second task.

Taken together, these results suggest that more cues and a more complex functional relationship make shifting from a rule-based strategy to another strategy more likely, while at the same time strengthen the preference for similarity-based strategies. As a consequence, participants' strategy choices are less predictable the more complex the task structure is.

Discussion

Study 1 examined whether providing scarce task feedback in categorization invites similarity-based processing and how the number of dimensions and the functional relation between cues and criterion influence reliance upon similarity-based strategies in judgment and categorization. We found that in both categorization and judgment, the OLIN task was best described by a cue abstraction model. Once more dimensions had to be integrated reliance on exemplar models increased with most people choosing an exemplar model in the MMULT task. These results replicate findings in judgment that judgments are better described by a similarity-based exemplar model than a cue abstraction strategy in tasks requiring the multiplicative combination of several cues (Hoffmann et al., 2013a; Juslin et al., 2008; Karlsson et al., 2007). This increased reliance upon the better performing similarity-based strategy indicates that task feedback helped adapting the cognitive process to task demands (Juslin, Olsson et al., 2003; Rieskamp & Otto, 2006). In addition, we showed that also the number of cues affects categorization and judgment strategies. The requirement to integrate more cues increased reliance on a similarity-based strategy suggesting that simple rules are more easily learnt than complex rules. This result resonates well with recent research showing that people abandon effortful cue abstraction processes more often under cognitive load (Filoteo et al., 2010; Hoffmann et al., 2013a; Zeithamova & Maddox, 2006). In sum, these results match well with the idea that the relative accuracy and effort of the strategies play an important role in strategy selection (Beach & Mitchell, 1978; Payne et al., 1993).

In contrast to our hypotheses we did not replicate the finding that more people relied on a cue abstraction strategy in categorization than in judgment in the MLIN task. There are

several potential explanations for this. In the first place, the rule-based cue abstraction strategy we used allows more complex rules than are mostly considered in categorization. In categorization rule-based processes are frequently restricted to conjunctive and disjunctive rules involving one or two dimensions. If two or more dimensions have to be integrated, for instance when learning optimal linear or nonlinear decision bounds, it is assumed that people rely on procedural learning (Ashby & Maddox, 2005; Ashby & O'Brien, 2005). In contrast, the judgment literature assumes that linear, additive relationships can be likewise learned by a rule-based cue abstraction strategy, drawing the line for rule abstraction between linear relationships and nonlinear relationships. Recent findings support the latter view: For instance, participants' explicit ratings of cue importance are highly correlated with cue weights derived from fitting a linear, additive model to linear tasks suggesting that people possess insight into the rule abstraction process (Lagnado, et al., 2006). Likewise, how well people learn rule-based as well as information-integration categorizations is associated with working memory capacity (Lewandowsky et al., 2012). Working memory capacity further predicts how accurate people make rule-based judgments, whereas similarity-based judgments rely more heavily on episodic memory (Hoffmann et al., 2013b). Furthermore, we did not find evidence that the majority of participants relied on a nonlinear bound in the MMULT categorization task[7] — what would have been expected if people indeed learned the optimal decision bound via a procedural learning process. Another reason why not more people relied on a cue abstraction strategy in categorization than in judgment could be that our task involved continuous instead of binary cues. Continuous cue values make it easier to abstract the direction of the relationship between a cue and the criterion and thereby facilitate the abstraction of cue weights (Newell et al., 2009), a factor that reliably enhances reliance on rule-based strategies (Platzer & Bröder, 2012; von Helversen et al., 2013).

With regard to the question if people approach judgment and categorization tasks similarly we found that the number of cues as well as the functional relation promoted

similarity-based processes in both judgment and categorization. However, more cues and a more complex function made participants shift more often between cognitive strategies from the first to the second task. One reason for this is possibly that more complex tasks make it more difficult to find the best way to solve the task so that people choose cognitive strategies more inconsistently. Alternatively, it has been proposed that individual preferences for learning based upon rules or exemplars are more pronounced in tasks that do not strongly favor one solution (McDaniel et al., 2013). People may learn over time how accurate and effortful it is to rely on rule-based strategies in comparison to exemplar-based strategies (Rieskamp & Otto, 2006). Consequently, people may build up stable tendencies for rule-based or similarity-based learning that may be related to stable personal characteristics like memory abilities (McDaniel et al., 2013). In study 2, we investigate if increasing the effort associated with exemplar memory reduces the demand characteristics of the task environment and hence makes preferences for rule-based and exemplar-based learning more prevalent.

Study 2

In study 2, we boost how effortful relying upon similarity-based strategies is by introducing a multidimensional quadratic (MQUAD) task structure in which the criterion is quadratic function of the cues. MQUAD judgment tasks are particularly hard to learn for two reasons: First, linear rules cannot be abstracted successfully so that they do not lead to good performance in training (Olsson et al., 2006). Second, an exemplar strategy can, in principle, learn to solve the task. However, the same criterion value is associated with multiple, but dissimilar exemplars making it more difficult to use exemplar memory (Olsson et al., 2006). Consequently, neither similarity-based processes nor rule-based processes yield to good performance early in training and people are supposed to drop back to the default, but useless rule abstraction process (Karlsson et al., 2008; Olsson et al., 2006). Indeed, people only solve quadratic judgment tasks if they are explicitly instructed to remember single instances (Olsson et al., 2006).

Interestingly, people can, however, master MQUAD categorization and one-dimensional quadratic function learning tasks (Ashby & Maddox, 1992; Ashby, Waldron, Lee, & Berkman, 2001; Busemeyer et al., 1997; Pachur & Olsson, 2012). One reason why people may still be able to solve MQUAD categorization tasks is possibly that it is easier to store only two different categories that are associated with the exemplars. In this vein, it has been found that also learning in MQUAD tasks that require categorizing exemplars into four different categories is significantly impaired (Ashby et al., 2001). Similar to Olsson et al. (2006), Ashby et al. (2001) concluded that people default to suboptimal linear decision rules. In sum, both studies suggest that people may shift to a large extent from rule-based strategies in judgment to similarity-based strategies in categorization.

However, MQUAD tasks may also foster reliance on personal preferences for rule-based or exemplar-based learning. For instance, McDaniel et al. (2013) found that people stick to their preferred learning strategy in linear V-shaped function learning tasks that are structurally most similar to quadratic judgment tasks. These learning preferences, in turn, transferred to how well people learned abstract categorization tasks. Hence, it is also possible that a MQUAD task again increases reliance on similar processes in judgment and categorization.

In sum, we expected to replicate the key finding from previous studies that people learn successfully to solve MQUAD categorization tasks, whereas learning should suffer more in MQUAD judgment tasks. Second, we tested if people still follow an exemplar-based strategy in MQUAD categorization tasks, but default to a rule-based strategy in MQUAD judgment tasks. To test these predictions, our participants solved a categorization and a judgment task with the same MQUAD task structure.

Method.

Participants.

Thirty-two participants (25 females, $M_{Age} = 26.5$, $SD_{Age} = 10.7$) were recruited from the University of Basel. Participants received course credit or a participation fee (20 CHF per hour) for participating in the experiment. In addition, they could earn a bonus of 3 CHF in each task and had the opportunity to win one of two Amazon vouchers (worth 25 CHF each).

Design and material.

We used the same cover stories and pictures as in Study 1. In the MQUAD task, the judgment criterion was a quadratic function of the cues:

$$y_{MQUAD} = 0.83 \left[4(c_1 - 2.5)^2 + 3(c_2 - 2.5)^2 + 2(c_3 - 2.5)^2 + (c_4 - 2.5)^2 - 2.5 \right] \quad (14)$$

where c_1 to c_4 are the cue values ranging from 0 to 5. According to their cue weights, c_1 reflects the most important cue and c_4 the least important one. Subtracting 2.5 from each cue centered the cue values on their mean. Consequently, high and low cue values are associated with higher criterion values, whereas intermediate cue values were associated with lower criterion values. The categories for the binary categorization task were created by performing a median split on the judgment criterion. This median split generates a category structure with a spherical category boundary. Accordingly, the less similar an exemplar is to the prototypical exemplar with intermediate cue values, the more likely it is that the exemplar belongs to a different category than the prototype.

To select a training set and a validation set, we generated again 1000 training sets with 25 training items and selected a training set that a) could not well be solved by one- or two dimensional rules in the categorization task and that b) was fitted worse by a (log-) linear regression than by an exemplar model. In a second step, we generated 100 validation sets consisting of 15 validation items and finally selected a validation set for which the cognitive models made strongly diverging predictions. The final training and validation sets are depicted in Table 6 and 7, respectively.

Procedure.

The procedure followed closely the procedure used in Study 1. Participants solved both a MQUAD judgment and categorization task. Task order as well as assignment of cover stories to the tasks was counterbalanced.

Like in Study 1, participants learned to predict the criterion values or the categories for the 25 training items over 10 training blocks. In the test phase, participants judged or categorized 15 validation items four times without getting any feedback. We encouraged participants to achieve a high performance by incentivizing their answers like in Study 1. In addition, participants could earn a bonus of 3 CHF if they reached more than 80% of the points in the last training block in the categorization task or more than 55% of the points in the last training block in the judgment task. The relaxed learning criterion in the judgment task corresponds approximately to a RMSD below 10 and accordingly participants reaching the learning criterion should outperform a linear model and a baseline model by 2 RMSD ($RMSD_{Baseline} = 12$, $RMSD_{Linear} = 11.8$).

Results

Performance in the categorization task.

Overall, participants solved the MQUAD categorization task worse than the MLIN or MMULT categorization tasks in Study 1 (see Table 3 for categorization and judgment performance). Although performance in the training phase dropped, still 25 participants (78.1%) reached a performance better than a baseline model predicting 44% of errors in the training phase.[8] Similarly, participants made more errors on validation items in the test phase than in study 1.

Performance in the judgment task.

Like in the categorization task, performance dropped in the judgment task compared to the MLIN and MMULT judgment tasks in Study 1. In both the training and the test phase participants made on average less accurate judgments than in Study 1.[9] However, still 19 participants (59.4%) outperformed a baseline model in the last training block. Because of the

different response scales it is difficult to compare judgment to categorization performance

based upon the RMSD (Mata et al., 2012). Therefore, we used the number of participants

classified as outperforming or falling behind the baseline model in the last training block of

the categorization and the judgment task. Overall, a marginal smaller amount of participants

fared better than the baseline model in the judgment task compared to the categorization task,

$b = 0.89$, $SE = .46$, Wald $\chi^2(1) = 3.80$, $p = .051$.

Model fits and deviances.

At the end of training, the exemplar model described participants' categorizations best

(see Table 4 for BICs, deviances, and strategy classifications), outperforming the baseline

model ($z = 3.14$, $p = .002$) and the cue abstraction model ($z = 4.43$, $p < .001$).[10] In the test

phase, the baseline model could not be distinguished from the cue abstraction model ($z = 1.22$,

$p = .224$). The exemplar model, however, predicted participants' categorization in the test

phase more accurately than the baseline model ($z = 3.14$, $p = .002$) or the cue abstraction

model ($z = 2.77$, $p = .006$).

In the judgment task, the BICs did not discriminate between the cognitive models (see

Table 5 for BICs, deviances, and strategy classifications). The BIC for the baseline model did

not differ from BICs for the cue abstraction model ($z = 0.654$, $p = .513$) or the exemplar

model ($z = 0.505$, $p = .614$). Neither could the cue abstraction model and the exemplar model

be distinguished ($z = 0.299$, $p = .765$). In the test phase, only the cue abstraction model made

more accurate predictions than the baseline model ($z = 3.09$, $p = .002$). The exemplar model

be discriminated from the baseline model ($z = 0.08$, $p = .940$), nor from the cue abstraction

model ($z = 0.75$, $p = .454$).

Strategy classification.

To analyze judgment and categorization strategies more closely and to assess how the

type of task may change the cognitive process, we classified participants to the baseline, the

cue abstraction, and the exemplar model. In categorization, this classification provided

stronger evidence for the exemplar model (see Figure 3). In judgment, however, a similar amount of participants was classified to the exemplar model and the cue abstraction model.

To test how the type of task (judgment or categorization) affected the cognitive process, we again performed an ordinal, logistic regression on strategy classification with task feedback as the independent variable. Overall, task feedback did not affect strategy classification, $b = 0.64$, $SE = .51$, Wald $\chi^2(1) = 1.52$, $p = .217$. This result suggests that although the judgment task was harder to learn, not more people tend to rely on similarity-based processes in the categorization than in the judgment task.

Matching processes in judgment and categorization.

Like in Study 1, we analyzed shifting behavior by classifying participants as following a cue abstraction model in the first and the second task, as following an exemplar model in both tasks or as shifting between processes. Descriptively, the majority of participants shifted between cognitive processes from one task to the other (21 participants). Four participants were best described by the cue abstraction model in both tasks and seven participants by the exemplar model. Figure 2 depicts the conditional probability of following a cue abstraction model (an exemplar model) in the second task given that participants followed a cue abstraction model (an exemplar model) in the first task they solved. Overall, $p(\text{Exemplar}_{\text{Second}}|\text{Exemplar}_{\text{First}})$ was higher than .5 indicating that participants tended to stay with the exemplar model in the second task. Interestingly, however, $p(\text{CAM}_{\text{Second}}|\text{CAM}_{\text{First}})$ was rather low suggesting that participants abandoned a cue abstraction strategy in the second task and shifted more to similarity-based strategies in the second task.

Discussion

Study 2 investigated how increasing the difficulty to execute similarity-based strategies affects performance and cognitive strategies in categorization and judgment. Matching previous research we found that MQUAD judgment tasks are particularly hard to learn in terms of judgment accuracy, whereas categorization accuracy shows smaller

decrements (Ashby & Maddox, 1992; Ashby et al., 2001; Pachur & Olsson, 2012). More participants could outperform a baseline model in the categorization task than in the judgment task. These results are in line with research showing that increasing the number of categories harms categorization performance in MQUAD tasks (Ashby et al., 2001).

Mimicking results from Study 1, however, the type of task did not change cognitive processes. Although, descriptively, a few more participants were classified to the exemplar model in the categorization task than in the judgment task, there was no effect of the type of task. Finally, participants were not simply stuck with an inefficient cue abstraction process in judgment. Indeed, analyzing shifting behavior suggested a practice effect: People were more likely to abandon cue abstraction, if they already experienced how difficult cue abstraction was in a first task. Possibly, people transferred knowledge of the task structure to the new task and deliberatively chose in the second task to rely upon a similarity-based strategy already early in training (Olsson et al., 2006). Indeed, only 11 participants followed the same strategy in both tasks suggesting that individual preferences are overridden by task feedback, even if these tasks make it at first difficult to figure out how to solve the task best.

General Discussion

The distinction between similarity and rules is core to many areas of cognitive science (Hahn & Chater, 1998; Pothos, 2005; Sloman, 1996), but little research has linked similarity-based and rule-based processes across different domains like judgment and categorization. We contributed to integrating judgment and categorization research by studying how the number of cues and the functional relationship between cue and criterion shape cognitive strategies in judgment and categorization. Specifically, we suggested that the functional relationship restricts the accuracy that can be achieved by relying upon rules or similarity. Binary task feedback in categorization and the number of cues, however, increase the effort associated with abstracting rules (Study 1) or with storing single exemplars (Study 2). Finally,

we examined to what extent categorization processes match the processes people rely on when making judgments.

Across both studies, we found that functions that increasingly deviate from linear relationships promote similarity-based processing in both judgment and categorization. Matching research from judgment (Hoffmann et al., 2013a; Juslin et al., 2008; Karlsson et al., 2007), this result highlights that categorization and judgment behavior is highly adapted to the task demands with the relative accuracy of rule-based and similarity-based strategies as one key determinant of strategy shifts. Beyond that, these results emphasize that the functional relationship also drives shifts between rule-based and similarity-based processes in categorization — a factor that has been rarely studied in categorization or led to ambiguous results (Maddox & Ashby 1993; McKinley & Nosofsky, 1995, 1996).

In line with past research, we also found that more cues make it more difficult to abstract explicit rules and enhance similarity-based strategies in categorization (Ashby et al., 2002; Filoteo et al., 2010; Maddox & Ashby, 2004; Zeithamova & Maddox, 2006) and judgment (Karelaia & Hogarth, 2008). This result suggests that increasing the effort necessary for rule abstraction further forces participants to adopt more similarity-based processes. However, in contrast to studies using binary cues (Juslin, Olsson et al., 2003; Mata et al., 2012; von Helversen et al., 2010, 2013), both studies did not find evidence that reduced task feedback in categorization invites more similarity-based processes than continuous feedback in judgment. One reason why scarce feedback may have less impact on processing is that the continuous cues we used implicitly convey knowledge about the cue directions (Newell et al., 2009) and hence may foster cue abstraction processes also more strongly in categorization (Platzer & Bröder, 2013; von Helversen et al., 2013). Apparently, continuous cue information can trigger cue abstraction and compensate for less informative feedback.

As one of the first studies, we directly investigate the extent to which people adopt similar strategies in the domains of judgment and categorization. Striking is the fact that the

more cues the task involved and the more complex the functional relationship between cues and criterion was, the more people tended to abandon rule-based strategies from the first to the second task. As a result, participants also shifted more between rule-based and similarity-based processes. One reason why these inconsistencies may arise is that task complexity makes it more difficult to detect the adequate solution to the task. Indeed, in the multidimensional quadratic task a huge percentage of participants at first seemed to fall back to rule abstraction (cue abstraction: 47%; exemplar: 34%). However, in the second task they experienced fewer problems in detecting the underlying task structure and shifted more to an exemplar-based process (cue abstraction: 31%; exemplar: 59%). In addition, this result also raises concerns against the idea that individual preferences for learning rule-based or similarity-based become stronger the less determined the task environment is (McDaniel et al., 2013).

In our study, however, we also found that still the majority of participants in Study 1 were classified as following the same process to judge and categorize objects suggesting that judgment and categorization processes build upon common principles. In the MLIN task, for instance, 50% of the participants were best described with a cue abstraction model in both tasks. While judgment research has argued that such cue abstraction models describe explicit cue abstraction processes, categorization research has often claimed that people learn linear boundaries implicitly (Ashby & Maddox, 2005; Ashby & O'Brien, 2005). These formed linear boundaries are difficult to verbalize and people, hence, do not build up any explicit task knowledge (Ashby & Maddox, 2005). Our result, however, resonates well with current findings showing that how well people solve rule-based categorization tasks is highly correlated with accuracy in information integration categorizations (Lewandowky, 2011) and that people adopting task-appropriate strategies in rule-based categorizations are also more likely to adopt the appropriate strategy in information integration (Lewandowsky et al., 2012).

In line with these findings, our results move explicit cue abstraction processes assumed in judgment closer to implicit information integration processes in categorization.

On a theoretical level, our study matches well with the idea that people can both rely on similarity-based and rule-based processes (Erickson & Kruschke, 1998; Juslin, Olsson et al., 2003; Nosofsky et al., 1994; von Helversen & Rieskamp, 2008). While our study conceptualized the interaction of rules and similarity as shifts between cognitive strategies, it is also likely that people base judgment and categorizations simultaneously on rules and similarity by blending these cognitive processes (Hahn, Prat-Sala, Pothos, & Brumby, 2010; von Helversen, Herzog, & Rieskamp, 2013). Future computational accounts may further exploit how rules and similarity interact by disentangling shifting and blending accounts.

Taken together, our study suggests that people approach complex tasks by relying more on similarity, whereas scarce feedback does not further alter how people make judgments. Complex tasks, however, also pose a challenge by rendering the identification of task-appropriate strategies more difficult. Studying how people deal with a range of cognitive tasks may thus help to identify the conditions systematically triggering rules and similarity and to explore or to limit the generality of those two accounts.

Footnotes

1. Information integration models (or decision bound models) assume that people learn to separate categories by implicitly forming linear or quadratic categories boundaries (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). A linear decision bound corresponds mathematically to our instantiation of cue abstraction categorization models. In contrast to cue abstraction models, decision bound models suggest that decision rules are learnt implicitly via procedural learning and cannot be verbalized (Ashby & O'Brien, 2005; Ashby & Maddox, 2005).

2. The order of the tasks (categorization or judgment task first) and the cover story (bugs or Sonics) did not affect how well people learned the task, $U = 1101$, $p = .706$, and $U = 1032.5$, $p = .377$, respectively.

3. The cover story (bugs or Sonics) did not affect judgment accuracy ($U = 1132$, $p = .883$), but order of the tasks (judgment or categorization first) had a marginal effect on judgment accuracy ($U = 898.5$, $p = .062$).

4. In addition we also fitted an exemplar model with four attention weights. This model did not outperform the predictions of an exemplar model with one parameter in the categorization task, OLIN: -2LL = 66.53 (SD = 78.04), MLIN: -2LL = 93.60 (50.32), MMULT: -2LL = 83.23 (64.29). Likewise, the exemplar model with four attention weights only generalized better in the OLIN judgment task than the exemplar model with one parameter, OLIN: -2LL = 20.08 ($SD = 185.00$), MLIN: -2LL = 135.92 ($SD = 17.72$), MMULT: -2LL = 130.12 ($SD = 16.18$). Furthermore, the linear model generalized better than the exemplar model with four attention weights in almost all task, except the MMULT judgment task.

5. We also tested two-dimensional conjunctive and disjunctive categorization rules. Overall, these rules did not describe a large number of participants best: OLIN: 3.1% (1 participant), MLIN: 15.6% (5 participants), MMULT: 15.6% (5 participants).

6. To not overweigh tiny differences in model predictions, likelihood values could not exceed 100 or fall below .001. Similarly, the fitted standard deviations had to exceed .001.

7. Only 8 participants (25%) were best described when fitting equation 3 to participants' categorizations in the MMULT task and predicting categorizations for validaion items. The cue abstraction model still predicted 13 participants (41 %) best and the exemplar model 11 participants (34 %).

8. Categorization performance in the last training block was neither affected by order, $F(1,29) = 2.15$, $p = .153$, nor the cover story, $F(1,29) = 0.14$, $p = .716$.

9. The order of the tasks did not affect judgment performance in the last training block, $F(1,29) = 0.79$, $p = .382$. Participants were slightly better at judging bugs (RMSD = 10.8, SD = 2.4) than Sonics in the last training block (RMSD = 13.1, SD = 3.0), $F(1,29) = 5.74$, p = .023.

10. Again, an exemplar model with four parameters did not make more accurate predictions than an exemplar model with one parameter in the categorization task (-2LL = 88.92, SD = 62.69) and the judgment task (-2LL = 157.12, SD = 11.27).

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A

neuropsychological theory of multiple systems in category learning. *Psychological

Review, 105*, 442–481. doi:10.1037/0033-295X.105.3.442

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of

multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory,

and Cognition, 14*, 35–53. doi:10.1037/0278-7393.14.1.33

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization:

Contrasting novice and experienced performance. *Journal of Experimental

Psychology: Human Perception and Performance, 18*, 50–71. doi:10.1037/0096-

1523.18.1.50

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of

Psychology*, *56*, 149–178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback

training in rule-based and information-integration category learning. *Memory &

Cognition, 30*, 666–677.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems.

Trends in Cognitive Sciences, 9, 83–89. doi:10.1016/j.tics.2004.12.003

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human

categorization and identification. *Journal of Experimental Psychology: General, 130,*

77-96. doi: 10.1037//0096-3445.130.1.77

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision

strategies. *The Academy of Management Review*, *3*, 439–449.

Busemeyer, J. R., Byun, E., DeLosh, E. L., & McDaniel, M. A. (1997). Learning functional

relations based on experience with input-output pairs by humans and artificial neural

networks. In Lamberts K. & D. Shanks (Eds.), *Concepts and Categories* (pp. 405–437). Cambridge: MIT Press.

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*, 171–189. doi:10.1006/jmps.1999.1282

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: the sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 968–986.

Einhorn, H. J. (1971). Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance, 6*, 1-27.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107–140. doi:10.1037/0096-3445.127.2.107

Filoteo, J. V., Lauritzen, J. S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science, 21*, 415–423. doi:10.1177/0956797610362646

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition, 65*, 197–230. doi:10.1016/S0010-0277(97)00044-9

Hahn, U., Prat-Sala, M., Pothos, E. M., & Brumby, D. P. (2010). Exemplar similarity and rule application. *Cognition, 114*, 1–18. doi:10.1016/j.cognition.2009.08.011

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013a). Deliberation's blindsight: How cognitive load can improve judgments. *Psychological Science, 26*, 869-879. doi: 10.1177/0956797612463581

Hoffmann, J A., von Helversen, B., & Rieskamp, J. (2013b). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. Manuscript submitted for publication.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 924–941.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*, 259–298. doi:10.1016/j.cognition.2007.02.003

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*, 133–156. doi:10.1037/0096-3445.132.1.133

Kämmer, J. E., Gaissmaier, W., & Czienskowski, U. (2013). The environment matters: Comparing individuals and dyads in their adaptive use of decision strategies. *Judgment and Decision Making, 8*, 299-329.

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*, 404–426. doi:10.1037/0033-2909.134.3.404

Karlsson, L., Juslin, P., & Olsson, H. (2007). Adaptive changes between cue abstraction and exemplar memory in a multiple-cue judgment task with continuous cues. *Psychonomic Bulletin & Review*, *14*, 1140–1146.

Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making : Contingent, not automatic, strategy shifts? *Judgment and Decision Making*, *3*, 244–260.

Kaufmann, E., & Athanasou, J. A. (2009). A meta-analysis of judgment achievement as defined by the lens model equation. *Swiss Journal of Psychology, 68*, 99–112. doi:10.1024/1421-0185.68.2.99

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in

    multiple-cue learning. *Journal of Experimental Psychology: General, 135,* 162–183.

    doi:10.1037/0096-3445.135.2.162

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual

    differences and modeling. *Journal of Experimental Psychology: Learning, Memory,*

    and Cognition, 37, 720–738. doi:10.1037/a0022639

Lewandowsky, S., Yang, L. X., Newell, B. R., & Kalish, M. L. (2012). Working memory

    does not dissociate between different perceptual categorization tasks. *Journal of*

    Experimental Psychology: Learning, Memory and Cognition, 38, 881–904. doi:

    10.1037/a0027298

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of

    categorization. *Perception & Psychophysics, 53*, 49–70. doi:10.3758/BF03211715

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based

    systems of perceptual category learning. *Behavioural Processes, 66*, 309–332.

    doi:10.1016/j.beproc.2004.03.011

Mata, R., Helversen, B. von, Karlsson, L., & Cüpper, L. (2012). Adult age differences in

    categorization and multiple-cue judgment. *Developmental Psychology*, *48*, 1188–

    1201. doi:10.1037/a0026084

McDaniel, M. A., & Busemeyer, J. R. (2005). The conceptual basis of function learning and

    extrapolation: Comparison of rule-based and associative-based models. *Psychonomic*

    Bulletin & Review, 12, 24–42. doi:10.3758/BF03196347

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2013). Individual differences in

    learning and transfer: Stable tendencies for learning exemplars versus abstracting

    rules. *Journal of Experimental Psychology: General.* Advance online publication.

    doi:10.1037/a0032963

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 128–148.

McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 294–317.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238. doi:10.1037/0033-295X.85.3.207

Newell, B. R., Weston, N. J., Tunney, R. J., & Shanks, D. R. (2009). The effectiveness of feedback in multiple-cue probability learning. *Quarterly Journal of Experimental Psychology, 62*, 890–908. doi:10.1080/17470210802351411

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700–708. doi:10.1037//0278-7393.14.4.700

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7*, 375–402.

Nosofsky, R. M., Little, D. R., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1–27. doi:10.1037/a0021330

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53–79. doi:10.1037/0033-295X.101.1.53

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesics and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247–255. doi:10.1111/1467-9280.00051

Olsson, A.-C., Enkvist, T., & Juslin, P. (2006). Go with the flow: How to master a nonlinear

multiple-cue judgment task. *Journal of Experimental Psychology: Learning,*

Memory, and Cognition, 32, 1371–1384. doi: 10.1037/0278- 7393.32.6.1371

Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy

selection in decision making. *Cognitive Psychology*, *65*, 207–240.

doi:10.1016/j.cogpsych.2012.03.003

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New

York, NY: Cambridge University Press.

Platzer, C., & Bröder, A. (2013). When the rule is ruled out: Exemplars and rules in decisions

from memory. *Journal of Behavioral Decision Making*, *26*, 429–441. doi:

10.1002/bdm.1776

Pothos, E. M. (2005). The rules versus similarity distinction. *The Behavioral and Brain*

Sciences, *28*, 1–14.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies.

Journal of Experimental Psychology: General, 135, 207–236. doi:10.1037/0096-

3445.135.2.207

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of*

Experimental Psychology: General, *133*, 63–82. doi:10.1037/0096-3445.133.1.63

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–

464.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological*

Bulletin, 119, 3–22. doi:10.1037/0033-2909.119.1.3

von Helversen, B., Herzog, S. M., & Rieskamp, J. (2013). Haunted by a Doppelgänger:

Irrelevant facial similarity affects rule-based judgments. *Experimental Psychology,*

29, 1-11.

von Helversen, B., Karlsson, L., Mata, R., & Wilke, A. (2013). Why does cue polarity

information provide benefits in inference problems? The role of strategy selection

and knowledge of cue importance. *Acta Psychologica, 144*, 73–82.

doi:10.1016/j.actpsy.2013.05.007

von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond

looks? From similarity-based to cue abstraction processes in multiple-cue judgment.

Developmental Psychology, 46, 220–229. doi:10.1037/a0016690

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of

quantitative estimation. *Journal of Experimental Psychology: General,* 137, 73–96.

doi:10.1037/0096-3445.137.1.73

von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-based

and exemplar-based processes compared. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 35, 867–889. doi:10.1037/a0015501

Wills, A. J., & Pothos, E. M. (2011). On the adequacy of current empirical evaluations of

formal models of categorization. *Psychological Bulletin, 138*, 102-125.

doi:10.1037/a0025715

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category

learning. *Memory & Cognition, 34*, 387–398. doi:10.3758/BF03193416

Table 1

Training Set for Study 1. Judgment Criteria and Categorizations were Derived from Equation 1 (MLIN), Equation 2 (OLIN), and Equation 3 (MMULT).

Cues				Judgment			Categorization		
Cue 1	Cue 2	Cue 3	Cue 4	MMULT	MLIN	OLIN	MMULT	MLIN	OLIN
2	1	0	3	2	14	0	0	0	0
1	4	1	4	5	22	10	0	0	0
0	3	1	2	2	13	10	0	0	0
0	2	3	0	1	12	30	0	0	1
5	5	4	0	29	43	40	1	1	1
0	4	5	4	12	26	50	1	1	1
2	4	3	0	9	26	30	1	1	1
1	4	3	5	13	27	30	1	1	1
1	0	2	4	1	12	20	0	0	0
1	0	0	2	1	6	0	0	0	0
5	3	3	5	21	40	30	1	1	1
1	1	5	5	7	22	50	1	0	1
1	2	0	5	2	15	0	0	0	0
5	5	0	1	4	36	0	0	1	0
0	4	3	1	4	19	30	0	0	1
4	2	1	3	6	27	10	1	1	0
0	5	2	3	6	22	20	1	0	0
5	5	2	4	22	43	20	1	1	0
5	1	3	4	9	33	30	1	1	1
4	0	2	4	3	24	20	0	0	0

1	4	1	5	6	23	10	1	0	0
3	0	5	5	3	27	50	0	1	1
0	2	5	0	2	16	50	0	0	1
1	5	2	4	10	27	20	1	1	0
3	4	5	5	30	39	50	1	1	1

Note. OLIN = One-dimensional, linear task; MLIN = Multidimensional, linear task; MMULT

= Multidimensional, multiplicative task

Table 2

Validation Set for Study 1. Judgment Criteria and Categorizations were Derived from Equation 1 (MLIN), Equation 2 (OLIN), and Equation 3 (MMULT).

Cues				Judgment			Categorization		
Cue 1	Cue 2	Cue 3	Cue 4	MMULT	MLIN	OLIN	MMULT	MLIN	OLIN
3	5	1	4	10	33	10	1	1	0
3	4	4	3	21	35	40	1	1	1
5	0	3	4	4	30	30	0	1	1
3	4	2	5	14	33	20	1	1	0
5	0	5	5	4	35	50	0	1	1
3	2	0	2	2	20	0	0	0	0
2	3	4	0	9	25	40	1	.5	1
4	5	4	5	36	44	40	1	1	1
5	0	5	3	4	33	50	0	1	1
4	3	0	1	3	26	0	0	1	0
2	1	2	0	3	15	20	0	0	0
2	5	2	3	12	30	20	1	1	0
4	0	0	2	2	18	0	0	0	0
4	1	1	1	4	22	10	0	0	0
3	3	3	5	15	32	30	1	1	1

Notes: OLIN = One-dimensional, linear task; MLIN = Multidimensional, linear task;

MMULT = Multidimensional, multiplicative task

Table 3

Performance in the OLIN, MLIN, and MMULT Categorization and Judgment Tasks in Study 1 and in the MQUAD Categorization and Judgment task in Study 2 (Standard Deviations in Parenthesis)

	Task condition			
	OLIN	MLIN	MMULT	MQUAD
Categorization task				
% errors Training	3.8 (8.7)	22.5 (9.1)	23.4 (12.9)	29.3 (15.5)
% errors Test	3.5 (8.3)	24.0 (11.1)	21.8 (13.1)	35.2 (18.4)
Judgment task				
RMSD Training	4.2 (8.0)	6.7 (3.1)	5.4 (2.1)	11.9 (2.9)
RMSD Test	3.4 (6.2)	5.8 (1.5)	5.5 (1.9)	14.2 (2.6)

Note. OLIN = One-dimensional, linear task; MLIN = Multidimensional, linear task; MMULT = Multidimensional, multiplicative task; MQUAD = Multidimensional, quadratic task; RMSD = Root mean square deviation

Table 4

Model Fits During Training and Test in the OLIN, MLIN, and MMULT Categorization Tasks in Study 1and in the MQUAD Categorization Task in Study 2 (Standard Deviations in Parenthesis)

	Model		
	Baseline	Cue abstraction	Exemplar
	OLIN		
BIC Training	103.64 (0.49)	35.29 (24.13)	25.08 (24.89)
Deviance Test	83.13 (0.70)	18.10 (22.66)	54.22 (28.76)
Classification (n)	0	27	5
	MLIN		
BIC Training	102.31 (2.22)	69.62 (20.39)	72.25 (17.44)
Deviance Test	84.91 (6.84)	82.10 (54.87)	71.38 (16.34)
Classification (n)	2	21	9
	MMULT		
BIC Training	101.39 (4.68)	79.39 (13.70)	74.33 (22.89)
Deviance Test	83.00 (3.07)	64.89 (23.90)	64.50 (11.01)
Classification (n)	0	17	15
	MQUAD		
BIC Training	101.59 (2.70)	111.08 (11.33)	87.64 (20.18)
Deviance Test	82.82 (3.84)	84.03 (10.54)	72.91 (14.79)
Classification (n)	5	9	18

Note. OLIN = One-dimensional, linear task; MLIN = Multidimensional, linear task; MMULT = Multidimensional, multiplicative task; MQUAD = Multidimensional, quadratic task; BIC = Bayesian Information Criterion

Table 5

Model Fits During Training and Test in the OLIN, MLIN, and MMULT Judgment tasks in

Study 1and in the MQUAD Judgment Task in Study 2 (Standard Deviations in Parenthesis)

	Model		
	Baseline	Cue abstraction	Exemplar
	OLIN		
BIC Training	1021.5 (26.2)	-300.2 (1127.8)	-396.5 (1255.2)
Deviance Test	193.7 (7.3)	-37.0 (175.4)	62.6 (106.8)
Classification (n)	2	30	0
	MLIN		
BIC Training	852.4 (34.1)	697.16 (90.3)	705.6 (98.0)
Deviance Test	157.2 (8.7)	130.3 (15.3)	133.6 (15.6)
Classification (n)	0	19	13
	MMULT		
BIC Training	772.0 (50.5)	680.0 (67.9)	628.46 (95.8)
Deviance Test	146.3 (12.3)	130.4 (17.9)	123.0 (14.7)
Classification (n)	1	13	18
	MQUAD		
BIC Training	871.22 (60.84)	871.13 (68.48)	859.44 (72.97)
Deviance Test	158.16 (10.26)	156.21 (11.60)	157.38 (12.32)
Classification (n)	4	16	12

Note. OLIN = One-dimensional, linear task; MLIN = Multidimensional, linear task; MMULT = Multidimensional, multiplicative task; MQUAD = Multidimensional, quadratic task; BIC = Bayesian Information Criterion

Table 6

Training Set for Study 2. Judgment Criteria and Categorizations were Derived from Equation

15 for the Multidimensional Quadratic Task

Cue 1	Cue 2	Cue 3	Cue 4	Judgment	Categorization
0	3	1	4	25	2
0	5	0	5	50	2
0	5	3	2	35	2
1	0	0	0	37	2
1	1	2	4	13	1
1	3	4	3	10	1
1	5	2	5	27	2
2	0	0	0	30	2
2	1	4	5	13	1
2	2	1	4	5	1
2	3	0	2	10	1
2	3	1	5	8	1
2	3	5	2	10	1
2	5	2	0	20	1
2	5	3	0	20	1
3	0	4	3	18	1
3	1	2	0	10	1
3	3	5	1	12	1
3	5	3	5	20	1
4	0	0	1	33	2
4	3	0	4	18	1

5	0	5	1	47	2
5	2	1	2	23	2
5	2	4	5	28	2
5	4	5	4	37	2

Table 7

Validation Set for Study 2. Judgment Criteria and Categorizations were Derived from

Equation 15 for the Multidimensional Quadratic Task.

Cue 1	Cue 2	Cue 3	Cue 4	Judgment	Categorization
0	2	0	3	30	2
0	4	2	0	30	2
1	0	1	3	25	2
1	2	2	2	7	1
1	5	5	2	32	2
2	3	1	3	3	1
3	2	3	0	5	1
3	2	4	1	5	1
3	3	0	3	10	1
3	3	1	4	5	1
3	3	3	2	0	1
3	4	4	3	8	1
4	2	2	3	7	1
5	2	4	1	25	2
5	4	5	3	35	2

Figure 1. Classification of participants to the baseline, the cue abstraction, and the exemplar model depending on the type of task (categorization or judgment) in Study 1. OLIN = one-dimensional, linear task; MLIN = Multidimensional, linear task; MMULT = Multidimensional, multiplicative task.
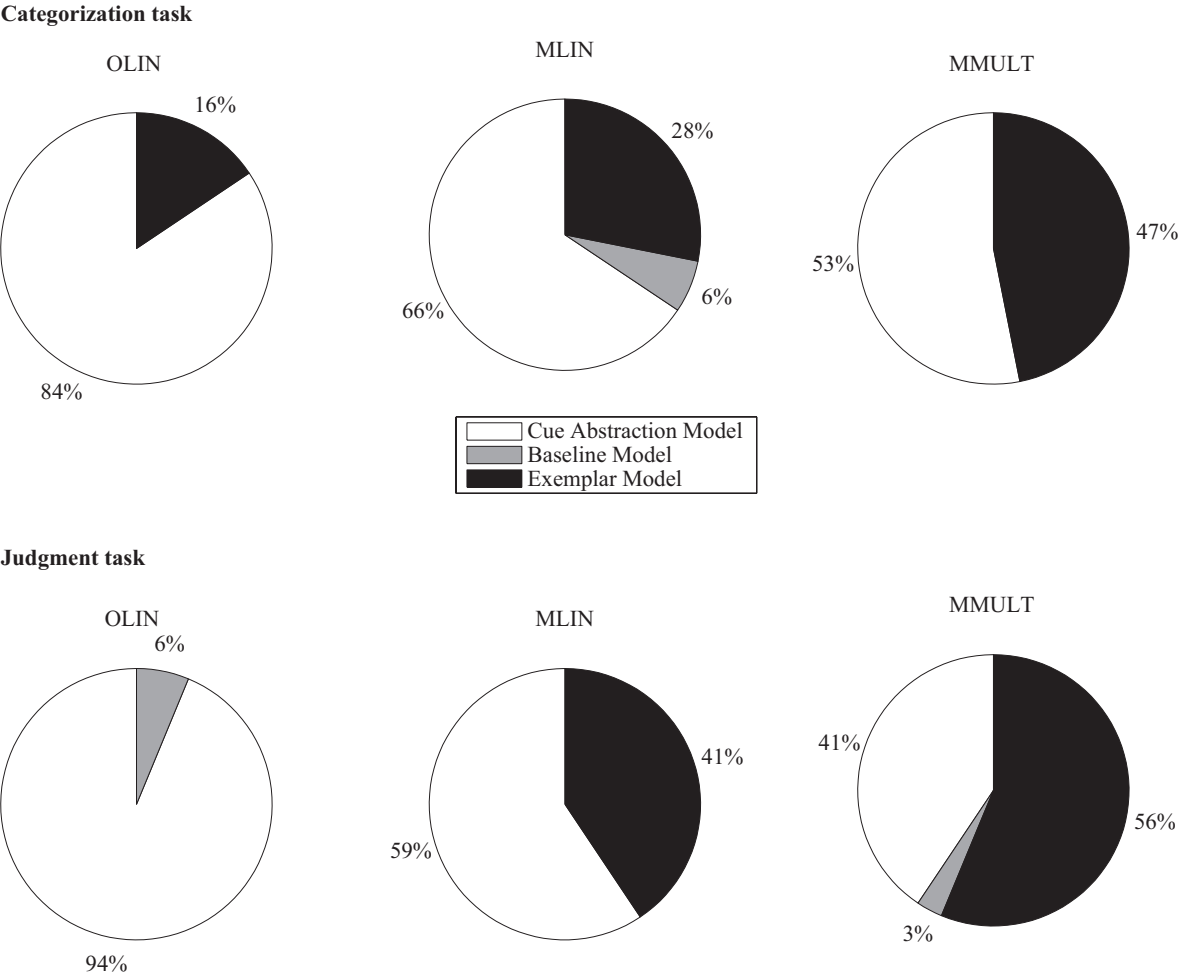
Figure 2. Conditional probabilities of following a cue abstraction model (CAM) or an exemplar model in the second task given that the participant followed a CAM or an exemplar model in the first task, respectively. Conditional probabilities are depicted for the OLIN (one-dimensional, linear), the MLIN (multidimensional, linear), and the MMULT (multidimensional, multiplicative) task from Study 1 as well as for the MQUAD (multidimensional, quadratic) task from Study 2.
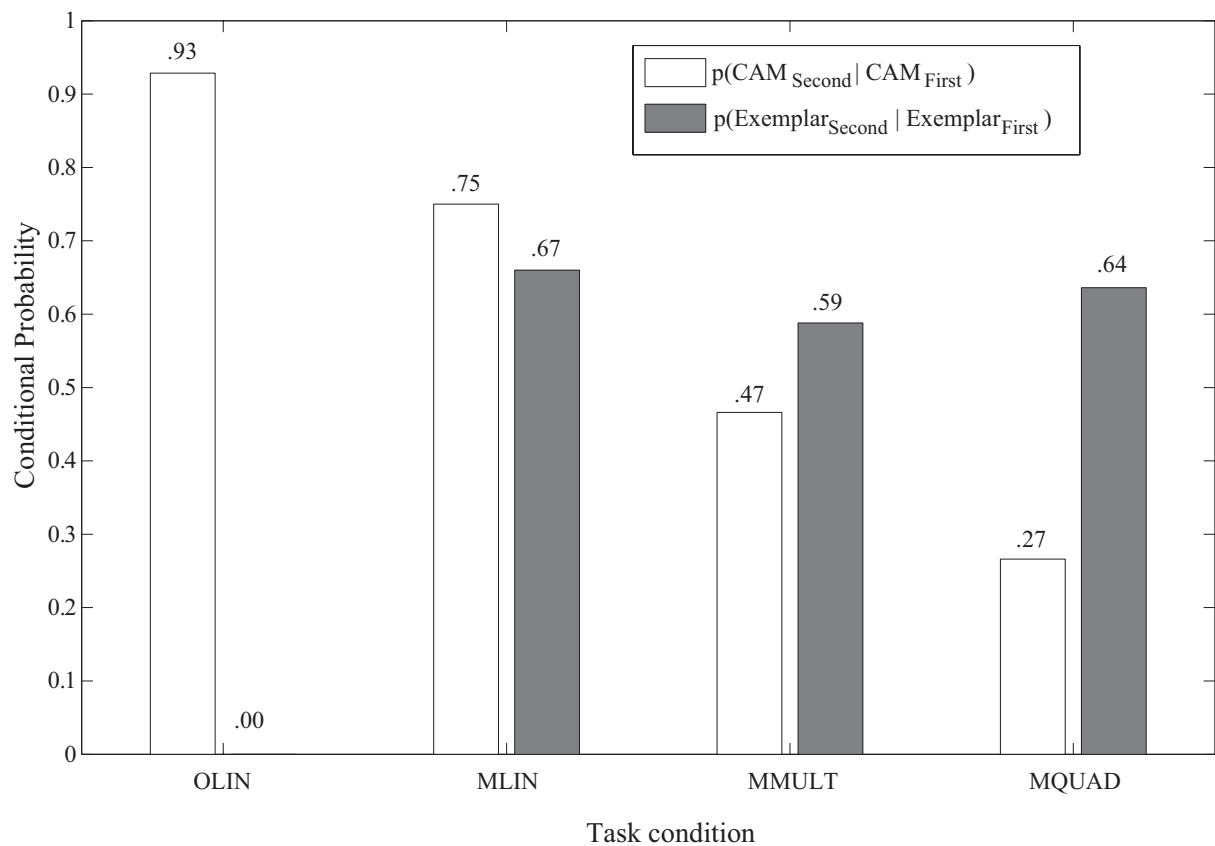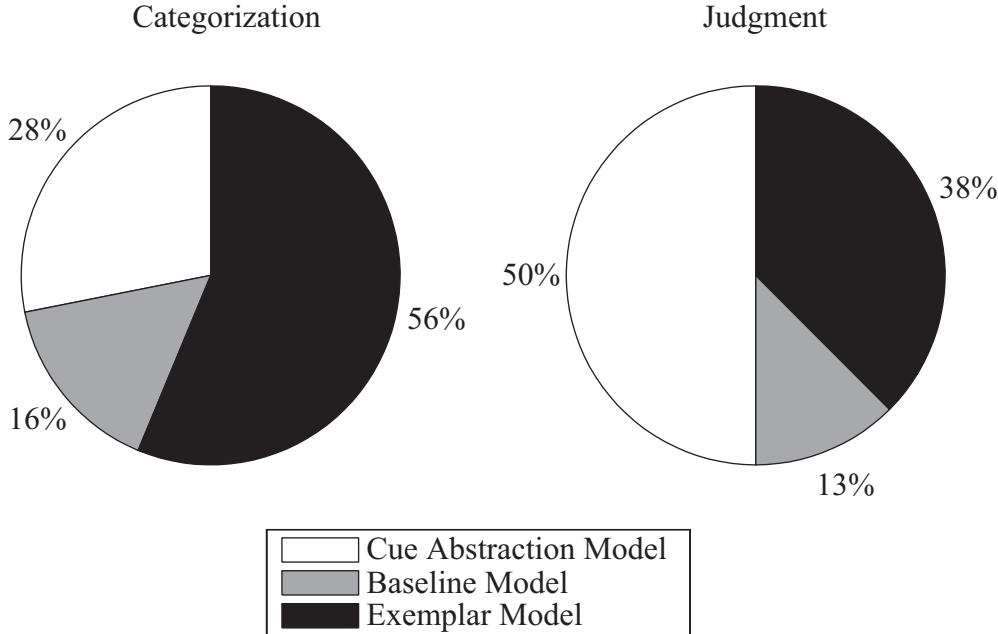
Figure 3. Classification of participants to the baseline, the cue abstraction, and the exemplar model depending on the type of task (judgment or categorization) in the multidimensional, quadratic task in Study 2.

Pillars of judgment: How memory abilities affect performance in rule-based and

exemplar-based judgments

Janina A. Hoffmann, Bettina von Helversen, and Jörg Rieskamp

University of Basel, Department of Psychology, Basel, Switzerland

Word count: 12,559

Abstract: 168

Author Note.   Janina A. Hoffmann, Bettina von Helversen, and Jörg Rieskamp,

Department of Psychology, University of Basel, Switzerland. This research was

supported by two grants from the Swiss National Science Foundation

(100014_130192/1 and 100014_146169/1).

Correspondence concerning this article should be addressed to Janina A. Hoffmann,

Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel,

Switzerland. Email: janina.hoffmann@unibas.ch.

Abstract

Making accurate judgments is an essential skill in everyday life. However,
although the relation of different memory abilities to categorization and judgment
processes has been hotly debated, the question is far from resolved. We contribute to
the solution by investigating how individual differences in memory abilities affect
judgment performance in two tasks that induce rule-based or exemplar-based
judgment strategies. In a study with 279 participants, we investigated how working
memory, episodic memory, and implicit memory affect judgment accuracy and
strategy use. As predicted, participants switched strategies between tasks.
Furthermore, structural equation modeling showed that the ability to solve rule-based
tasks was predicted by working memory, whereas episodic memory predicted
judgment accuracy in the exemplar-based task. We did not find evidence that
judgment accuracy was related to implicit memory. Last, the probability of choosing
an exemplar-based strategy was related to better episodic memory, but strategy
selection was unrelated to working memory capacity. In sum, our results suggest that
different memory abilities are essential for successfully adopting different judgment
strategies.

Keywords: Judgment; working memory; episodic memory; rule-based and
exemplar-based processes

"The only way to learn the rules of this Game of games is to take the usual prescribed

course, which requires many years." (Hermann Hesse)

In Hesse's fictitious country Castalia, one of the greatest honors is to be elected as a

Magister Ludi, the master of the glass bead game. This glass bead game integrates

knowledge from all the major scholarly disciplines — ranging from mathematics to

music to philosophy — by storing this academic knowledge in the form of game

symbols. During the game, these symbols are combined to form new ideas according

to the grammar of the game. A challenging glass bead play thus hinges on two

cornerstones of cognition: long-term memory and working memory. On the one hand,

a glass bead player needs to store knowledge in long-term memory and retrieve this

knowledge during the game. On the other hand, combining this knowledge requires

the ability to manipulate information while keeping it activated for a short time — one

key function of working memory.

Long-term memory and working memory are crucial for solving various tasks

in everyday life. When shopping, for example, it is necessary to remember the items

you intended to buy — a typical long-term memory task. Quickly summing up the

prices in your shopping basket, by contrast, places strong demands on working

memory. The ability to make accurate judgments may also hinge on basic memory

processes. To judge, for instance, the attractiveness of a job offer, people may recall

past work experiences from long-term memory. Alternatively, people may form an

initial judgment and repeatedly update this judgment by gathering information from

the job advertisement — a process that draws on key functions of working memory.

These examples clearly highlight that it is hardly possible to think of judgments

without considering memory abilities.

Indeed, the role of memory processes in making judgments cannot be overstated (Weber, Goldstein, & Barlas, 1995). Consequently, the interplay of long-term memory and working memory plays a major role in theories in categorization, judgment, and decision making (Ashby & O'Brien, 2005; Gigerenzer, Todd, & the ABC Research Group, 1999; Juslin, Karlsson, & Olsson, 2008; Marewski & Schooler, 2011). In particular the question of the degree to which different categorization and judgment strategies draw on distinct memory systems has animated a heated scientific debate (Ashby & O'Brien, 2005; Knowlton, 1999; Lewandowsky, 2011; Newell, Dunn, & Kalish, 2011; Nosofsky & Zaki, 1998; Smith, Patalano, & Jonides, 1998). In this vein, a growing body of research investigating the role of working memory capacity has suggested that higher working memory capacity helps people make more accurate judgments and categorizations (Lewandowsky, 2011; Weaver & Stewart, 2012). However, the contribution of long-term memory has been largely ignored in empirical research (Ashby & O'Brien, 2005; Del Missier et al., 2013; Tomlinson, Marewski, & Dougherty, 2011). Furthermore, we can think of no study that considered how various memory abilities interact with different categorization or judgment strategies.

Our goal was to fill this gap and shed light on which memory abilities underlie judgments. Specifically, we investigated how individual differences in working memory, episodic memory, and implicit memory interact with the judgment strategies people use. Focusing on two fundamental judgment strategies — rule-based and exemplar-based strategies (Erickson & Kruschke, 1998; Juslin, Olsson, & Olsson, 2003; von Helversen & Rieskamp, 2008, 2009) — we examined how memory abilities influence the selection and execution of these judgment strategies and, ultimately, judgment performance.

We first provide an overview about memory abilities and the strategies underlying human judgments. We then explore theoretically how judgment strategies are grounded in memory processes and how memory abilities encourage the selection of different judgment strategies. Finally, we report an individual difference study examining how memory abilities influence judgment accuracy and strategy use.

Memory Abilities

Memory refers to people's ability to store information. Memory research has drawn a major distinction between long-term memory and working memory. While long-term memory stores information for a long time period from minutes to years, working memory serves the purpose of manipulating information and maintaining this information in a highly active state for a short time (Atkinson & Shiffrin, 1968). Recent theories often understand working memory as consisting of activated representations in long-term memory (Oberauer, 2009; Unsworth & Engle, 2007). Indeed, evidence from individual difference studies suggests that working memory correlates with performance in long-term memory tasks (Del Missier et al., 2013; Mogle, Lovett, Stawski, & Sliwinski, 2008; Unsworth, 2010). Specifically, working memory may control encoding into and strategic retrieval from long-term memory (Baddeley, Lewis, Eldridge, & Thomson, 1984; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Rosen & Engle, 1997; Unsworth, Brewer, & Spillers, 2013).

Furthermore, memory research has drawn a prominent distinction between implicit and episodic long-term memory (we use the term *episodic memory* here to refer to explicit long-term memory for specific events). Whereas episodic memory measures reflect conscious recollection of facts or episodes, in implicit memory tests previous experiences facilitate performance, but these performance effects do not require conscious recollection of past experiences (Roediger, 1990; Squire & Zola, 1996). Countless studies have shown dissociations between implicit and episodic

memory tests and these dissociations have often been taken as evidence for two

distinct memory systems (Squire & Zola, 1996). For instance, correlation studies

showed that implicit memory measures, such as word stem completion, are not

correlated with episodic memory measures, such as cued recall (Bruss & Mitchell,

2009; Fleischman, Wilson, Gabrieli, Bienias, & Bennett, 2004; Perruchet & Beaveux,

1989). At the same time, however, the idea that there exist distinct episodic and

implicit memory systems has been repeatedly challenged (e.g., Berry, Shanks,

Speekenbrink, & Henson, 2012; Dew & Cabeza, 2011; Roediger, 1990). Recently, for

instance, Berry et al. (2012) suggested that one single process model accommodates

performance differences between recognition and implicit repetition priming tests. In

addition, several studies raised methodological concerns about the reliability of

implicit memory measures (Buchner & Brandt, 2003; Buchner & Wippich, 2000;

Meier & Perrig, 2000). All this considered, it is still an open question if episodic and

implicit memory can best be understood as two distinct memory systems.

Judgment Strategies

People make judgments every day ranging from estimating the probability of

rainfall to judging the attractiveness of a job. Making such judgments requires

inferring a continuous criterion, for instance, job attractiveness, from a number of

critical attributes of this object (i.e., the cues), such as the yearly income or the task

demands. People may rely on two different types of judgment strategies: rule-based

and exemplar-based (Erickson & Kruschke, 1998; Juslin et al., 2003; von Helversen &

Rieskamp, 2008, 2009).

Rule-based strategies assume that people form hypotheses about the

relationship between the cues and the criterion and apply this knowledge to make a

judgment (Brehmer, 1994; Juslin et al., 2008). Rule-based judgment strategies have

been predominantly captured with a linear, additive model (Cooksey, 1996) or a cue

abstraction model (Juslin et al., 2003). Linear models describe people's judgments in a

variety of tasks ranging from personal selection (Graves & Karren, 1992) to medical

diagnoses (Wigton, 1996) and have been found to match people's explicitly stated

judgment rules (Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Lagnado, Newell, Kahan,

& Shanks, 2006). Based on the lens model (Brunswik, 1956), the linear model

assumes that people explicitly abstract a weight for each cue and then combine the

weighted cue values in an additive fashion (Einhorn et al., 1979, Juslin et al., 2003).

For instance, when judging the attractiveness of a job offer, people first determine how

much they value income and the variety of task demands. Then they weight the yearly

income and task demands of the job by their respective importance and combine this

knowledge by adding the weighted cue values.

Exemplar-based judgment strategies, by contrast, rely on the retrieval of past

experiences from exemplar memory. For instance, when judging the attractiveness of a

new job, people may think about past jobs they have held. Exemplar-based strategies

assume that previously encountered objects are stored in memory along with their

criterion values (Juslin et al., 2003, 2008). To judge the new object (the probe),

previously encountered objects (exemplars) are retrieved from memory. For instance,

when judging the attractiveness of a job offer, a job applicant may recall previous

work experiences. The more similar a retrieved exemplar is to the probe, the more it

influences the final judgment. Accordingly, if a job applicant worked in a job with

similar task demands, he might just recall how much he liked his former job to rate the

attractiveness of the new job offer. Thus, exemplar-based strategies imply that people

store concrete instances without abstracting any knowledge and engage in an

associative similarity-based process during retrieval.

In sum, rule-based and exemplar-based strategies differ in their assumptions

about the cognitive processes underlying judgments (Hahn & Chater, 1998; Juslin et

al., 2003). Whereas rule-based strategies use abstracted knowledge about the world to reason about new instances, similarity-based or exemplar-based strategies rely on the similarity to past instances. Research suggests that both strategies are frequently used, with strategy selection depending on task characteristics (Juslin et al., 2003, 2008; Karlsson, Juslin, & Olsson, 2007; Platzer & Bröder, 2013; von Helversen, Karlsson, Mata, & Wilke, 2013; von Helversen & Rieskamp, 2009) and individual differences (Mata, von Helversen, Karlsson, & Cüpper, 2012; von Helversen, Mata, & Olsson, 2010): Specifically, people rely more on cue abstraction strategies in linear judgment tasks where the criterion is a linear additive function of the cues but shift to exemplar-based strategies in multiplicative judgment tasks where the criterion is a nonlinear function of the cues (Hoffmann, von Helversen, & Rieskamp, 2013a; Juslin et al., 2008). This shift presumably takes place because the cue abstraction strategy does not allow accurate judgments in nonlinear environments (Juslin et al., 2008; von Helversen & Rieskamp, 2009). In the following section, we review theoretical and empirical work on how the cognitive processes underlying rule-based and exemplar-based strategies map onto different memory abilities.

Linking Judgment Strategies and Memory Abilities

In general, memory abilities can limit two different aspects of strategy use: strategy execution and strategy selection (Beach & Mitchell, 1978; Lemaire & Siegler, 1995; Mata, Pachur, et al., 2012). First, memory abilities can influence strategy execution, the ability to execute a strategy correctly. Better episodic memory, for instance, can enhance exemplar retrieval from memory and thus lead to more accurate exemplar-based judgments. Second, memory abilities can influence strategy selection by either fostering the ability to choose the more accurate strategy or boosting the preference for a single strategy (Beach & Mitchell, 1978). We first address the question of how the execution of rule-based and exemplar-based strategies are related

to working memory, episodic memory, and implicit memory and thereafter address the question of strategy selection.

The Influence of Memory Abilities on Strategy Execution

       Rule-based strategies. Solving a rule-based categorization or judgment task has often been equated with logical reasoning (Ashby & O'Brien, 2005) or problem solving (Juslin et al., 2008). Like reasoning or problem-solving tasks, rule-based strategies such as cue abstraction are thought to involve a serial, controlled hypothesis-testing process and, in turn, working memory (Ashby & O'Brien, 2005; Brehmer, 1994; Juslin et al., 2003, 2008). Specifically, working memory may be required by two aspects of the rule-based judgment process: rule abstraction and rule execution.

       Relying on cue abstraction requires abstracting the cue weight, the weight that should be given to a specific cue. One way this can be achieved is by comparing two objects, relating the difference in judgment criteria to the difference in cue values, and then updating the cue weights accordingly (Juslin et al., 2008; Pachur & Olsson, 2012). This comparison process likely taxes working memory, because it involves storing information about the two judgment objects for a short time and actively manipulating this information, key functions of working memory (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974). Overall, recent research supports this idea, showing that learning rules hinges on working memory. Learning simple, one-dimensional categorization rules, for instance, is impaired by a concurrent verbal task (Filoteo, Lauritzen, & Maddox, 2010; Zeithamova & Maddox, 2006, 2007). In a similar vein, cognitive load studies in judgment have suggested that people abandon cue abstraction strategies more frequently under cognitive load than without load (Hoffmann et al., 2013a). Finally, learning a judgment task is easier if the sequence reduces working memory demands by facilitating a direct comparison of cue values

and judgment criteria (Helsdingen, Van Gog, & Van Merriënboer, 2011; Juslin et al., 2008).

Not only learning a rule, but also applying a rule may involve working memory processes, such as mental updating and inhibition (Miyake et al., 2000; Oberauer, 2009). When making a judgment people may start with an initial estimate that is updated with each new piece of evidence (Hogarth & Einhorn, 1992; Juslin et al., 2008) — a process that requires keeping the past estimate in mind and manipulating it mentally. Furthermore, rule application requires inhibiting information, because people need to focus attention on the relevant cues and ignore those that are irrelevant. In line with this idea, Del Missier et al. (2013) found that correctly applying decision rules was related to working memory capacity. Specifically, rule application involved inhibiting irrelevant information and updating information in working memory (Del Missier, Mäntylä, & Bruine de Bruin, 2010, 2012).

Long-term memory may be less important for making rule-based judgments, compared to working memory. Once a rule has been established, only the cue weights need to be retrieved from long-term memory (Bruner, Goodnow, & Austin, 1956). Previously encountered objects, in contrast, can be forgotten (von Helversen & Rieskamp, 2008), so that episodic memory should have a negligible influence on rule execution.

Exemplar-based strategies. Exemplar-based strategies assume that judgments are based on the similarity to previously encountered exemplars (Juslin et al., 2003; Medin & Schaffer, 1978; Nosofsky, 1988), suggesting that executing exemplar-based strategies should be linked to episodic memory (Hintzman, 1986, 1988; Nosofsky, 1988). Basically, two major types of episodic memory processes may contribute to successfully adopting exemplar-based strategies: encoding into and retrieval from episodic memory (Estes, 1986; Shiffrin & Atkinson, 1969).

Before any information can be recalled from memory, it is necessary to form a memory representation (i.e., to encode) and store this information (Estes, 1986). Like episodic trace models of human episodic memory, for instance, MINERVA 2 (Hintzman, 1984, 1986), most exemplar-based models assume that exemplars are encoded in separate memory traces, storing each presentation of an exemplar in a single trace (Estes, 1986; Nosofsky, 1988). Accordingly, the more often an object is presented, the more often it is encoded and the more likely is its subsequent retrieval. Likewise, elaboration, adding information to the memory trace, or spacing exemplar presentations across time intervals can deepen encoding (Brown & Craik, 2000; Martin, 1968). Beyond storing the exemplars in episodic memory, successfully adopting an exemplar-based strategy also requires accurately retrieving the stored exemplars from episodic memory. Retrieval may fail because the probe's features — serving as retrieval cues — do not activate memory traces for stored exemplars or past exemplars can no longer be discriminated (Medin & Schaffer, 1978).

Although theoretical accounts suggest strong links between episodic memory and exemplar-based strategies, empirical evidence for the relationship is still scarce (Ashby & O'Brien, 2005). Nevertheless, it has been shown that the instruction to learn all exemplars by heart helps learning in judgment tasks solvable by exemplar strategies (Olsson, Enkvist, & Juslin, 2006). Likewise, if single exemplars have to be memorized to solve a categorization task, these exemplars are recognized more easily in a subsequent recognition test (Davis, Love, & Preston, 2012; Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004). In contrast, if people cannot identify past exemplars, they are less inclined to follow exemplar-based strategies (Rouder & Ratcliff, 2004). Furthermore, similar to spacing effects in memory (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), spacing exemplar repetitions helps when solving exemplar-based tasks (McDaniel, Fadler, & Pashler, 2013).

Neuropsychological work has challenged the view that similarity-based category learning depends solely on episodic memory (Knowlton, 1999; Smith, 2008). The multiple-systems view (Ashby & O'Brian, 2005; Smith & Grossman, 2008) proposes instead that some category structures are learned implicitly. Specifically, it has been argued that implicit memory underlies prototype distortion tasks in which new items have to be categorized based on the similarity to a prototype extracted from previously encountered exemplars. For instance, Knowlton and Squire (1993) found that amnesiac patients classified new items with the same accuracy as a healthy control group but were less accurate at recognizing patterns they had seen before. Similar dissociations between amnesiac patients and a control group have been found in implicit memory research. While amnesiac patients are severely impaired in recognizing or recalling previously studied words, they do not show performance deficits in implicit word completion tests (Graf, Squire, & Mandler, 1984). Accordingly, Smith and Grossman (2008, p. 259) concluded that "similarity-based categorization can be based on either explicit or implicit memory." Likewise, Juslin et al. (2008) indicated that exemplar-based strategies might be driven by different representations, including perceptual traces and semantic memory structures.

However, proponents of exemplar-based accounts have rejected the idea that exemplar-based strategies rely on an implicit memory system distinct from episodic memory (Nosofsky & Zaki, 1998). They have argued that dissociations between categorization and recognition can be easily explained by a single exemplar model. Thus, it is still an open question if implicit memory is related to exemplar-based judgments.

Besides implicit and episodic memory, working memory could also be helpful for learning in exemplar-based judgment tasks. Lewandowsky (2011), for instance, argued that every recollection-based long-term memory task should be related to

working memory capacity. Underpinning his argument, working memory has been found to support encoding and retrieval processes in episodic memory (Baddeley et al., 1984; Craik et al., 1996; Rosen & Engle, 1997; Unsworth et al., 2013). Retrieving past exemplars may also involve a deliberative search process in long-term memory (Juslin et al., 2008; Karlsson, Juslin, & Olsson, 2008). Indeed, research suggests that working memory load not only harms rule-based strategies but also disturbs retrieving past exemplars when judging new objects (Juslin et al., 2008). Furthermore, learning to solve rule-based and exemplar-based categorization tasks is facilitated by high working memory capacity (Craig & Lewandowsky, 2012; Lewandowsky, 2011; Lewandowsky, Yang, Newell, & Kalish, 2012). Therefore, working memory capacity should — in general — promote executing exemplar-based judgment strategies. However, if working memory promotes exemplar-based processing by enhancing episodic memory, episodic memory will serve as a mediator between working memory capacity and exemplar-based judgments, and hence, working memory capacity should lose importance for predicting exemplar-based judgments.

The Influence of Memory on Strategy Selection

Beyond influencing strategy execution, memory abilities could also influence which strategies people choose (Hoffmann et al., 2013a). The demands rule-based and exemplar-based strategies place on specific memory abilities can be conceptualized as costs and benefits of using a strategy. To choose a strategy, people may learn to trade off the benefits and costs associated with each strategy (Beach & Mitchell, 1978; Payne, Bettman, & Johnson, 1993; Rieskamp, 2006; Rieskamp & Otto, 2006). Hence, memory abilities could strengthen or weaken the preference for employing a specific strategy. In this vein, people with good episodic memory may take advantage of their skills and select an exemplar-based strategy more often, whereas people with bad episodic memory may avoid remembering past exemplars. In line with this idea, it has

been found that older adults avoid following an exemplar-based strategy — possibly because it places high demands on episodic memory (Mata, von Helversen et al., 2012). In the same way, high working memory capacity may facilitate using rules and thus encourage rule-based processing.

However, there is also good reason to believe that memory abilities differentially affect selecting a rule- or exemplar-based strategy. When learning to make judgments, people seem to start with a rule and only switch to an exemplar-based strategy if the rule fails (Juslin et al., 2008; Nosofsky, Palmeri, & McKinley, 1994). Accordingly, if rule-based strategies serve as a default option, memory abilities such as high working memory capacity may not be required to select a rule-based judgment strategy, but only to execute the rule-based strategy successfully.

Beyond influencing preferences for specific strategies, memory abilities could also influence the general ability to choose the strategies adaptively (Mata, Pachur et al., 2012). Consistently, Bröder (2003) found that more intelligent participants tended to select a strategy that ignores information when this strategy performs well. Similarly, people with higher working memory capacity do not simply prefer rule-based strategies in categorization; instead they seem to select the more appropriate strategy for the task at hand (Craig & Lewandowsky, 2012; Lewandowsky et al., 2012). Thus, people with high working memory capacity may not only apply rules more accurately but may also be faster in detecting when rules cannot properly solve exemplar-based judgment tasks, prompting a shift to exemplar-based strategies.

Predictions for judgment performance and strategy selection

To predict how memory abilities are related to judgment performance, it is necessary to take the judgment task into account. Research suggests that people prefer rule-based strategies in linear judgment tasks but switch to exemplar-based strategies in multiplicative judgment tasks (Hoffmann et al., 2013a; Juslin et al., 2008). Thus,

memory abilities should differentially affect judgment performance in linear and multiplicative tasks.

Specifically, low working memory capacity should harm the execution of cue abstraction strategies, because incorrect cue weights are learned or applying the learned rule is disrupted. In contrast, poor episodic memory should only marginally influence the execution of a cue abstraction strategy above working memory. Consequently, higher working memory capacity but not better episodic memory should be linked to more accurate judgments in linear, additive judgment tasks. Successfully executing an exemplar-based strategy, in contrast, hinges on encoding into and retrieval from episodic memory so that better episodic memory — and possibly implicit memory abilities — should improve judgment accuracy in multiplicative judgment tasks, whereas working memory should not affect judgment performance in a multiplicative task above episodic memory. Regarding strategy selection, working memory capacity may help people to detect and choose the more appropriate strategy in linear and multiplicative judgment tasks. Episodic memory, in contrast, may make it more likely that people rely on retrieval of past exemplars in multiplicative judgment tasks.

The Present Study

The current study examined how memory abilities relate to judgment performance in two different judgment tasks: a linear, additive judgment task and a multiplicative judgment task. Additionally, we measured working memory, episodic memory, and implicit memory with three different tests each. We selected the memory tests so that variance stemming from material or task-specific effects was reduced, allowing us to measure relatively pure latent abilities (Miyake et al., 2000). For this purpose, we used memory tests that included different types of material (verbal,

spatial, or numeric) and different types of tests (e.g. recognition, cued recall, and free recall for episodic memory).

Participants

Two hundred and seventy-nine participants (147 female, 132 male, M_{Age} = 24.0, SD_{Age} = 6.0) were recruited at the University of Basel. Participants received an hourly fee for their participation (20 Swiss francs, CHF, approx. U.S. $22) and could earn an additional bonus in the judgment tasks (M = 10.3 CHF, SD = 2.5 CHF). Overall, it took participants about 5 hr to complete the study, including a break of half an hour.

Automated Working Memory Span Tasks

Working memory span tasks were designed to measure both storage and processing of information in working memory (Redick et al., 2012), by letting participants process one set of stimuli while remembering another set of stimuli. For instance, in each trial of the operation span task, participants first see a simple equation. After they solve the equation and give the answer, they see the first letter that has to be remembered. Subsequently, another equation is presented and another letter has to be remembered, until the set size (the number of to-be-remembered letters) is reached. Finally, participants are asked to recall the letters in the order of their appearance. Trials with different set sizes are randomly interspersed, with each set size repeated three times.

We used three different span tasks that are often used in individual differences studies (Unsworth, Redick, Heitz, Broadway, & Engle, 2009; Unsworth, McMillan, Brewer, & Spillers, 2012): the reading span, the operation span, and the symmetry span. All span tasks were taken from Unsworth et al. (2009) and translated into German. We measured working memory capacity using the partial credit score as the

dependent variable (Conway et al., 2005). The partial credit score is the sum of all items recalled in the correct position over all trials.

Operation span. Participants were asked to solve mathematical equations while remembering letters. Set size varied from 3 to 7 so that partial credit scores could range from 0 to 75.

Reading span. In the reading span participants judged the plausibility of a sentence while remembering letters.[1] Set size varied from 3 to 7 so that partial credit scores could range from 0 to 75.

Symmetry span. Participants judged the symmetry of a chessboard picture while remembering the positions of squares in a 4×4 matrix. In each trial, participants first saw a chessboard picture and were asked to judge its symmetry. Afterward, one square in the 4×4 matrix was highlighted and participants were asked to remember its position. After the set size had been reached, participants recalled the positions of the squares by clicking on the squares in the order of their appearance. Set size varied from 2 to 5 so that partial credit scores could range from 0 to 42.

Episodic Memory Tasks

We measured episodic memory with three different tasks: a free recall task with pictures, a cued recall task with numbers, and a recognition test of verbs.

Picture free recall. We selected 20 pictures from a picture database (Rossion & Pourtois, 2004) that had high ratings on imagery and concreteness. Each picture was presented for 3 s on a computer screen and participants were asked to remember them. After a retention interval of 2 min participants recalled the pictures by naming them. Performance was measured as the percentage of correctly recalled pictures.

Cued number recall. We assessed cued number recall with a computerized version of the cued number recall task from the Berliner Intelligenzstruktur-Test Form 4 (BIS 4; Jäger, Süß, & Beauducel, 1997). Fifteen pairs of a two- and a three-digit

number were first presented for 10 s on the screen. After a retention interval of 2 min, participants saw the cued number pair as well as four, three-digit number distractors and had to indicate which three-digit number was initially presented together with the two-digit number. Performance was measured as the percentage of correctly recalled three-digit numbers.

Verb recognition. We selected 40 verbs with five to seven letters from the Hager and Hasselhorn database (1994), which is rated high on imagery and concreteness. Twenty verbs were assigned to a list of old items and 20 to a list of new items with the two lists matched on word length, imagery, and concreteness. In the study phase, participants learned the old verbs for 3 s each. After a retention interval of 2 min, participants indicated whether they recognized the 40 verbs from the study phase by classifying them as *old* or *new*. Performance was measured as the percentage of verbs correctly classified as old or new.

Implicit Memory Tasks

Previous studies have questioned the reliability of implicit memory measures (Buchner & Brandt, 2003; Buchner & Wippich, 2000; Meier & Perrig, 2000). To increase the reliability, we followed the suggestion of Buchner and Brandt (2003) and used performance tests that always had a correct solution (instead of association tests such as word stem completion). Our participants solved three different implicit memory tests: a speeded presentation test of line drawing, an identification test for sounds presented in noise, and an identification test for degraded nouns.

We measured performance in the implicit memory tasks as the difference in median reaction times between old and new items, including correct and incorrect answers. Negative reaction time differences indicate that participants responded faster to the old items than to the new items, showing an implicit memory effect, a facilitation of performance because of prior experience.

Speeded presentation of line drawings. The design of the speeded presentation task followed closely an experiment by Musen and Treisman (1990). We randomly created 500 line drawings. From these line drawings we excluded duplicates and drawings representing simple forms, such as arrows. From the remaining items we randomly selected 40 line drawings — 20 old and 20 new — for the implicit memory test with the restriction that they had at most two lines in common. To determine the presentation threshold we used 40 different line drawings from the remaining items. These line drawings had at least two lines that were different from all items used in the implicit memory test.

Using a threshold procedure we first determined the presentation length at which participants could correctly reproduce half of the line drawings. Starting with a presentation length of 400 frames (approx. 1200 ms), participants were asked to retrace the briefly presented line drawing on a mask that was composed of all lines possible in the line drawing. Participants were forced to draw all five lines. If they could not remember all the lines they were asked to guess. After each correct reproduction the presentation length decreased by 100 frames (300 ms). After each incorrect drawing, the presentation length increased by 100 frames. We decreased the step size to 10 frames (30 ms) after five turning points (the term turning point refers to a switch between decreases and increases in presentation length).

In the subsequent implicit learning phase, participants were asked to click as fast as possible on all lines of the 20 old items. Participants retraced all old items twice. After a 2-min retention interval, participants again completed a speeded reproduction task. The presentation length was set to the presentation length after the last trial of the threshold in the reproduction task. Participants were asked to redraw the briefly presented old and new line drawings. Performance was measured as the difference in median reaction times between old and new line drawings.

Identification of degraded nouns. Forty nouns with a high rating on imagery and concreteness and a length of five to seven letters were selected from the Hager and Hasselhorn (1994) database. Nouns that were highly similar to each other in spelling were excluded. The nouns were alphabetically sorted and 20 items with the same initials were randomly included in the old and new item list. To present the nouns in a degraded fashion, we superimposed an 8×2 chessboard mask over each noun. Nine of the 16 squares were randomly turned black, so that identification of the noun was made difficult.[2]

In the study phase, participants were asked to count the vowels in 20 nouns, with German umlauts counting as two vowels. Each noun was presented for 3 s on screen. After a retention interval of 2 min, participants were asked to correctly identify 40 degraded nouns by typing in the noun names. Half of the nouns were *old*; that is, they had already been presented in the study phase. Performance was measured as the difference in median reaction times between old and new degraded nouns.

Sound identification in noise. We selected 40 sounds from the Database for Environmental Sound Research and Application (Gygi & Shafiro, 2010) with a length between 0.55 and 3.54 s. All sounds were equalized for RMS (root mean squared) loudness, so that mean RMS loudness was 60 dB. For the sound identification task, the sounds were embedded in 5 s of white noise with a signal-to-noise ratio of -15 dB. Each sound started 0.5 s after stimulus onset.

In the study phase, participants were asked to indicate whether the 20 old sounds had a higher or lower pitch than their own voice. After a 2-min retention interval, participants listened to 20 old sounds from the study phase and 20 new sounds, all embedded in noise.[3] After each sound, participants were shown the name of the sound as well as the names of two other sounds that never appeared in the study

and had to indicate which of the sounds they had listened to. Performance was measured as the difference in median reaction times between old and new sounds.

Judgment Tasks

Participants solved both a linear and a multiplicative judgment task, taken from Hoffmann, von Helversen, and Rieskamp (2013b). In both tasks, participants had to judge a continuous criterion ranging from 0 to 50 based on four cues varying on a continuous scale from 0 to 5. In the linear judgment task, the criterion y was a linear, additive function of the cues:

$$y = 4c_1 + 3c_2 + 2c_3 + c_4, \tag{1}$$

where c_1 reflects the most important cue according to its cue weight. Each cue value varied between 0 and 5. In the multiplicative judgment task the function generating the criterion y included a multiplicative combination of the cues:

$$y = \frac{4c_1 + 3c_2 + 2c_3 + c_4 + 2c_1 c_2 c_3 + c_2 c_3 c_4}{8.5} \tag{2}$$

We used two different cover stories for the linear and the multiplicative multiple-cue judgment task. In the linear judgment task, participants judged whether a comic figure was a good or bad catcher of small creatures. In the multiplicative judgment task, participants estimated the toxicity of a bug. The stimuli for the two cover stories consisted of pictures of either bugs or comic figures. These bugs and comic figures varied on four cues. The bugs varied on the length of their legs, their antennae, and their wings, and the number of spots on their back. The comic figures had different sizes of ears and nose and a different number of hairs and stripes on their shirt. Table 1 illustrates the task structure: The cues could be used to predict the correct criterion value. The visual features were randomly assigned to the cues. The items were divided into a training set and a validation set. In the linear task, both sets could be better solved by applying a rule-based judgment strategy; in the

multiplicative task, however, an exemplar-based strategy should lead to a better

performance. Additionally, the rule-based and the exemplar-based strategy predicted

different responses on the validation items.

Both tasks consisted of a training phase and a test phase. During the training

phase, participants learned to estimate the criterion values for 25 training items from

the training set. In each trial, participants first saw a picture of a bug or a comic figure

and were asked to estimate its criterion value. Afterward they received feedback about

the correct value, their own estimate, and the points they had earned. The training

phase ended after 10 training blocks, each consisting of the 25 training items presented

in a random sequence. In the subsequent test phase, participants judged 15 new

validation items four times but did not receive any performance feedback.

To motivate participants to reach a high performance, participants could earn

points in every trial. The number of points they earned was a truncated quadratic

function of the deviation of their judgment j from the criterion y:

$$\text{Points} = 20 - \frac{(j - y)^2}{7.625} \tag{3}$$

At the end of the judgment tasks, the points earned were converted to a

monetary bonus (1,500 points = 1 CHF). In addition, participants earned a bonus of 3

CHF if they reached 80% of the points in the last training block (corresponding to a

root mean square deviation [RMSD] of less than 5.5 in both judgment tasks).

Filler Tasks

The filler tasks for the retention intervals were matched with the memory tests

so that they did not include the same stimulus material. All filler tasks were paper-and-

pencil versions of the tests. We used six mostly attention-based filler tasks: the d2 Test

(Brickenkamp, 2002), the underline "x," the letter series, the mark numbers divisible

by 7, and the number-symbol task from the BIS 4 (Jäger et al., 1997), as well as the

letter sets task from the Kit of Factor-Referenced Cognitive Tests (KIT; Ekstrom,

French, Harman, & Dermen, 1976). In the d2 attention test, for instance, participants

are asked to cross out all d's with two small dashes while ignoring all p's or d's with

more (or fewer) dashes (Brickenkamp, 2002).

Procedure

Participants solved all tasks on one day with a half-hour break between the two

sessions. The tasks were presented in the same order to each participant. In the first

session, participants first solved the linear judgment task. Afterward, they moved on to

the operation span, then solved the verb recognition (filler task: number-symbol test),

the sound identification in noise (filler task: letter series), and the picture free recall

task (filler task: underline x), and finally they completed the symmetry span.

The second session started with the multiplicative judgment task. Afterward,

participants completed the reading span, the degraded identification of nouns (filler

task: mark numbers divisible by 7), the cued number recall task (filler task: d2 Test),

and the speeded presentation of line drawings (filler task: letter sets).

Results

In a first step, we analyzed participants' average performance in the memory

tasks and judgment tasks (see Table 2 for descriptive statistics) and modeled

participants' judgment strategies. In a second step, we fitted a measurement model to

memory abilities and judgment performance separately. Next, we linked these two

measurement models, estimating a structural model that predicts judgment accuracy

by memory abilities. Finally, we investigated how strategy execution and strategy

selection in the judgment tasks influences the relationship between memory abilities

and judgment accuracy.

Performance Measures

Performance in the memory tasks. Performance in the working memory span tasks was comparable to normative data (Redick et al., 2012). Participants achieved a higher partial credit score in the operation and the reading span than in the symmetry span, indicating that they recalled more items in these tasks. In the episodic memory tasks, participants showed a higher recall rate in the recognition task than in the free recall or the cued recall task. In the implicit memory tasks, participants showed, on average, a higher implicit memory effect in the degraded presentation task than in the identification in noise task or the speeded presentation task. In the speeded presentation task, participants did not respond faster to the old items at all.

Performance in the judgment tasks. At first, we assessed how well participants learned to solve the judgment tasks. As an indicator of judgment performance, we calculated the RMSD between participants' judgments in the last training block and the correct criterion, with lower RMSDs indicating higher judgment accuracy. We used Wilcoxon z tests to compare performance in the judgment tasks, because the judgment data showed slight deviations from normality.

Overall, participants successfully learned to solve the judgment tasks. However, more participants earned a bonus in the multiplicative judgment task (81% of the participants) than in the linear judgment task (52% of the participants), $\chi^2(1) = 7.56, p = .006$. Also, participants judged the training items on average more accurately in the multiplicative judgment task than in the additive judgment task, Wilcoxon $z = 4.92, p < .001$.

Next, we investigated how well people could generalize their performance to new validation items in the test phase. Judgment performance for validation items was measured as the RSMD between the correct criterion and participants' mean judgment, that is, the judgment for each validation item averaged over the four presentations in the test phase. Judgment performance in the test phase was

comparable between the linear and the multiplicative judgment task (Wilcoxon $z = 1.46$, $p = .145$) but improved slightly compared to the training phase in both judgment tasks. This improvement was probably caused by a more restricted range of criterion values.

Modeling of Judgment Strategies

To investigate which judgment strategy participants relied on, we adopted a cognitive modeling approach. For each participant, we fitted a linear regression model (describing the rule-based strategy), an exemplar model (describing an exemplar-based strategy), and a baseline model (estimating participants' mean judgments) to participants' judgments in the last three blocks of the training phase and predicted participants' mean judgments for validation items by using the fitted parameter estimates (von Helversen & Rieskamp, 2008). This so-called generalization test possesses the advantage that it accounts for model complexity not only in terms of the number of free parameters but also in terms of their functional form (Busemeyer & Wang, 2000). We then compared the models based on the RMSD between model predictions and participants' judgments in the training phase and the test phase. We used Wilcoxon sign-ranked tests for these model comparisons because the RMSDs were not normally distributed.

Linear model. Linear regression models have been used to mathematically describe rule-based judgment strategies. In linear models, the importance of each cue for making a judgment is reflected in its cue weight; the higher the cue weights are, the more they influence the final judgment. The final criterion estimate $\hat{c}_p$ of an object p is the weighted sum of the cue values x_{pi}:

$$\hat{c}_p = k + \sum_{i=1}^{I} w_i \cdot x_{pi} \tag{4}$$

where w_i are the cue weights for each cue i and k is a constant intercept.

Exemplar model. To describe the exemplar-based strategy mathematically we used an exemplar model with one free sensitivity parameter (Juslin et al., 2003).[4] In exemplar models, the similarity $S(p,j)$ between the probe p and the exemplar j is an exponential decay function of the objects' distances d_{pj} (Nosofsky & Zaki, 1998):

$$S(p,j) = e^{-d_{pj}}. \tag{5}$$

This distance is determined by summing up the absolute differences between the cue values x_{pi} of the probe and the cue values x_{ji} of the exemplar on each cue i and then weighting this sum by the sensitivity parameter h.

$$d_{pj} = h\left(\sum_{i=1}^{I} |x_{pi} - x_{ji}|\right). \tag{6}$$

Correspondingly, the more closely the cue values of the probe and the exemplar match each other, the smaller the distance is between the objects. The sensitivity parameter expresses how strongly people discriminate among the stored exemplars. A sensitivity parameter close to 0 indicates no discrimination and a high sensitivity parameter indicates that people specifically remember each exemplar.

The criterion estimate $\hat{c}_p$ is then determined as the average sum of the similarities weighted by their corresponding criterion values c_j.

$$\hat{c}_p = \frac{\sum_{j=1}^{J} S(p,j) \cdot c_j}{\sum_{j=1}^{J} S(p,j)}. \tag{7}$$

Model fits. At the end of training, the baseline model did not provide a good description of participants' judgments in the linear and the multiplicative judgment task (see Table 3 for fit indices during training and test). In the linear judgment task, the linear model described participants' judgments overall better than the exemplar

model ($z = 14.5$, $p < .001$), whereas the linear model did not outperform the exemplar model in the multiplicative judgment task ($z = 1.5$, $p = .145$).

In the test phase, the linear model also accounted better for participants' judgments than the exemplar model in the linear judgment task ($z = 11.2$, $p < .001$). In contrast, the exemplar model made slightly more accurate predictions for participants' judgments than the linear model in the test phase of the multiplicative judgment task ($z = 4.8$, $p < .001$). Replicating the results from the training phase, the baseline model described participants' judgments worse than the linear model or the exemplar model in the linear judgment task (linear model: $z = 14.1$, $p < .001$; exemplar model: $z = 14.2$, $p < .001$) and the multiplicative judgment task (linear model: $z = 14.0$, $p < .001$; exemplar model: $z = 14.0$, $p < .001$).

Strategy classification. To further examine individual differences in strategy selection, we classified participants as selecting the strategy that led to the smallest RMSD between model predictions and participants' mean judgments. As shown in Figure 1, most participants adapted their judgment strategy to the judgment task. Whereas in the linear judgment task the majority of participants were best described by the linear model ($n_{Linear} = 220$, $n_{Exemplar} = 42$, $n_{Baseline} = 17$), in the multiplicative judgment task most participants were classified as following an exemplar model ($n_{Linear} = 99$, $n_{Exemplar} = 176$, $n_{Baseline} = 4$), $\chi^2(2) = 136.31$, $p = .001$. Indeed, half of the participants (50.2%) switched from a linear, rule-based strategy in the linear judgment task to an exemplar-based strategy in the multiplicative judgment task.

To capture how much participants relied on a cue abstraction or an exemplar-based strategy, we also fitted a strategy weight parameter to participants' judgments in the test phase, excluding participants best described by the baseline model (Hoffmann et al., 2013a). This strategy weight can take values between 0 and 1 and weights the predictions of the exemplar and linear models.

$$\hat{c}_p = w_s \cdot \hat{c}_{p,\text{Exemplar}} + (1 - w_s) \cdot \hat{c}_{p,\text{Linear}} \qquad (8)$$

A strategy weight above .5 indicates a higher probability for the exemplar model; a strategy weight below .5 indicates a higher probability for the linear model.

Because the strategy weights were not normally distributed, we calculated a one-sample Wilcoxon test. Indeed, the strategy weight was on average below .5 in the linear judgment task, $n = 262$, $M = .22$, $SD = .28$, skewness = 1.32, kurtosis = 0.81, $z = 11.3$, $p < .001$, whereas it was larger than .5 in the multiplicative judgment task, $n = 275$, $M = .60$, $SD = .38$, skewness = -.52, kurtosis = -1.26, $z = -3.9$, $p < .001$.

Taken together, our results underscore that participants' judgment processes were highly task sensitive (Juslin et al., 2008; Karlsson et al., 2007). While most participants relied on a rule-based judgment strategy in the linear judgment task, the majority adopted an exemplar-based strategy in the multiplicative judgment task.

Structural equation modeling

To understand how judgment performance is grounded in memory abilities, we followed a structural equation modeling approach. One particular strength of structural equation modeling is that it allows testing of theories about relations between theoretically well-defined latent constructs extracted from manifest indicators (Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007). In doing so, structural equation modeling corrects for task-specific variance, providing measurement-error-free estimates of the latent construct (Tomarken & Waller, 2005). A recommended approach (Anderson & Gerbing, 1988) is first to estimate the measurement model that relates the manifest indicators to the latent constructs and then to test the relations between the latent constructs based on theoretical assumptions.

Model fit is often evaluated based on several fit indices (Iacobucci, 2010; Kline, 2011) among them chi-square (χ^2), the standardized root mean square residual

(SRMR), the comparative fit index (CFI), and the root mean square error of

approximation (RMSEA). Because descriptive data indicated some deviations from

multivariate normality, we estimated all models using a maximum likelihood estimator

with robust standard errors (MLR) and Satorra–Bentler scaled χ^2 values (scaling

factor, SF) for χ^2 difference tests (Satorra & Bentler, 2001).

Measurement model for memory abilities. To establish construct validity,

we first estimated a measurement model for memory abilities from the memory data.

We hypothesized that working memory, episodic memory, and implicit memory

constitute three separate latent constructs, each described by three tests (episodic

memory: recognition, free recall, and cued recall; working memory: operation span,

reading span, and symmetry span; implicit memory: degraded presentation, speeded

presentation, and identification in noise). Although working memory and episodic

memory are typically positively correlated (Brewer & Unsworth, 2012), implicit

memory should be uncorrelated with episodic memory (Bruss & Mitchell, 2009) and

is probably uncorrelated with working memory, as well. Table 4 depicts the zero-order

correlations between all memory and judgment tasks. Because we could not estimate a

measurement model for implicit memory, we fixed all unstandardized factor loadings

for the implicit memory measures to 1 in all measurement models.

Indeed, as illustrated in Figure 2, a three-factor latent-variable model that

assumed working memory and episodic memory are correlated but independent from

implicit memory provided the best fit, $\chi^2(28) = 38.08$, SF $= 0.97$, $p = .097$, CFI $= .95$,

RMSEA $= .04$, SRMR $= .05$. Model fit was neither improved by adding a correlation

between implicit memory and working memory, $\chi^2(27) = 38.63$, SF $= 0.95$, $p = .069$,

CFI $= .94$, RMSEA $= .04$, SRMR $= .05$, nor by adding a correlation between implicit

memory and episodic memory, $\chi^2(27) = 38.06$, SF $= 0.97$, $p = .077$, CFI $= .94$,

RMSEA $= .04$, SRMR $= .04$. Furthermore, assuming that working memory, episodic

memory, and implicit memory are uncorrelated decreased model fit, $\chi^2(29) = 45.81$,

SF = 0.99, $p = .025$, CFI = .91, RMSEA = .05, SRMR = .06, $\Delta\chi^2(1) = 5.28$, $p = .022$.

These results replicate the key finding from previous individual difference

studies: that working memory and episodic memory are moderately correlated

(Brewer & Unsworth, 2012; Del Missier et al., 2013; Unsworth, 2010). In addition,

our results support the assumption that implicit memory is independent from episodic

memory (Bruss & Mitchell, 2009) and working memory.

Measurement model for judgment performance. The measurement model

for judgment performance was particularly interesting because — to our knowledge —

judgment research has not yet investigated if performance in linear and multiplicative

judgment tasks is task-specific or depends on a more general ability to learn

judgments. To measure judgment performance in both tasks, we used the RMSD

between participants' judgments and the correct criterion in each of the four test

blocks of the two tasks (see Table 3 for zero-order correlations). Judgment

performance in the linear judgment task was assumed to constitute one latent factor,

whereas judgment performance in the multiplicative task constituted the second latent

factor. We then compared three measurement models against each other, assuming (a)

that the latent factors are completely uncorrelated, (b) that the latent factors are

correlated, or (c) that the latent factors are identical; that is, performance over all test

blocks in the linear and the multiplicative judgment task can be described by one

latent factor.

As illustrated in Figure 3, a measurement model that assumed a correlation

between performance in the linear judgment task and performance in the

multiplicative judgment task provided the best fit, $\chi^2(19) = 21.87$, SF = 1.23, $p = .291$,

CFI = 1.00, RMSEA = .02, SRMR = .03, suggesting two moderately correlated latent

factors. Omitting the correlation between the latent factors did not harm model fit with

regard to CFI (0.99) and RMSEA (.05). However, the other two fit criteria yielded a different picture, $\chi^2(20) = 33.84$, SF = 1.24, $p = .027$, SRMR = .11, $\Delta\chi^2(1) = 10.29$, $p = .001$. A model that assumed a single latent factor for judgment performance was rejected by all fit criteria, $\chi^2(20) = 571.79$, SF = 1.15, $p < .001$, CFI = 0.53, RMSEA = .31, SRMR = .23.

Taken together, the small correlation between judgment accuracy in the linear and the multiplicative task yields some evidence that individual differences in judgment performance partly stem from a general ability to solve judgment problems. However, a huge amount of the individual differences in judgment performance were idiosyncratic to the multiplicative or the linear judgment task, suggesting that distinct processes may account for individual differences in task performance.

Predicting Judgment Performance With Memory Abilities

Do individual differences in memory abilities determine how well people solve different judgment tasks? We predicted that participants with higher working memory capacity should make more accurate judgments in the linear judgment task, whereas participants with better episodic memory skills should solve multiplicative judgment tasks more accurately. To test this hypothesis against competing ideas, we combined the measurement model for memory abilities with the measurement model for judgment performance into one structural model that assumes a path from working memory to judgment performance in the linear task and a path from episodic memory to judgment performance in the multiplicative task. We compared this model to three alternative models (1) a null model that assumes memory abilities do not predict judgment performance at all, (2) a model that assumes implicit memory further predicts performance in multiplicative judgment tasks, and (3) a full model that additionally assumes working memory predicts judgment performance in

multiplicative tasks and episodic memory predicts judgment performance in linear

tasks.

The hypothesized structural model captured the underlying covariance

structure very well, $\chi^2(117) = 110.71$, SF = 1.01, $p = .646$, CFI = 1.00, RMSEA = .00,

SRMR = .05, and better than the three alternative models: Assuming no relationship

between memory abilities and judgment performance decreased model fit

considerably, $\chi^2(119) = 149.79$, SF = 1.01, $p = .030$, CFI = 0.98, RMSEA = .03,

SRMR = .09, $\Delta\chi^2(2) = 35.22$, $p < .001$. Indeed, omitting the path from working

memory to judgment performance in the linear task decreased model fit, $\chi^2(118) =$

133.94, SF = 1.01, $p = .150$, CFI = 0.99, RMSEA = .02, SRMR = .08, $\Delta\chi^2(1) = 17.49$,

$p < .001$. Likewise, omitting the path from episodic memory to judgment performance

in the multiplicative task decreased model fit, $\chi^2(118) = 130.94$, SF = 1.01. $p = .197$,

CFI = 0.99, RMSEA = .02, SRMR = .07, $\Delta\chi^2(1) = 20.25$, $p < .001$. Also, including

implicit memory could not further explain performance differences in the

multiplicative task, $\chi^2(116) = 110.54$, SF = 1.01, $p = .626$, CFI = 1.00, RMSEA = .00,

SRMR = .05, $\Delta\chi^2(1) = 0.05$, $p = .823$. Finally, also the full model that assumed

working memory and episodic memory are both important for predicting judgment

performance in the linear and the multiplicative judgment task did not outperform the

hypothesized model, $\chi^2(115) = 107.02$, SF = 1.00, $p = .690$, CFI = 1.00, RMSEA =

.00, SRMR = .04, $\Delta\chi^2(2) = 3.41$, $p = .182$.

In line with our hypothesis, the resulting structural model (Figure 4) shows that

people with higher working memory capacity solved linear judgment tasks more

accurately than people with lower working-memory capacity, whereas people with

better episodic memory skills solved multiplicative judgment tasks better than people

with bad episodic memory abilities.[5] In the next step, we investigated if memory abilities also influence strategy selection.

Tracing the Path From Memory Abilities to Judgment Performance Through Judgment Strategies

      Strategy selection. In the Introduction we outlined that memory abilities may change strategy selection in two possible ways. On the one hand, working memory may make it more likely that people detect the task-appropriate judgment strategy faster; accordingly, working memory should predict strategy selection in the linear and the multiplicative task and strategy selection, in turn, predicts judgment accuracy. On the other hand, it is possible that an active strategy selection is only necessary for executing exemplar-based strategies. In this case, episodic memory may only predict strategy selection in the multiplicative task.

      To investigate how memory abilities affect strategy selection and, in turn, judgment accuracy, we relied on mediation analyses. If memory abilities influence judgment accuracy by altering the judgment strategy, then strategy selection should mediate the relationship between memory abilities and judgment performance. In doing so we compared a null model that assumed strategy selection does not mediate the relationship between memory abilities and judgment accuracy against different mediator models. Alternative models proposed that (a) strategy selection mediates the relationship between episodic memory and performance only in the multiplicative judgment task, (b) strategy selection in addition mediates the relationship between working memory and performance in the linear task, or (c) working memory additionally predicts strategy selection in the multiplicative task.

      To conduct theses analyses, we relied on the continuous strategy weight. Because the strategy weight indicates only how much participants relied on an exemplar-based strategy or a cue abstraction strategy, participants classified as

following a baseline model in the linear or the multiplicative task were coded as missing on that variable. To avoid excluding all their data, we used a full information maximum likelihood approach to estimate the structural model (Tomarken & Waller, 2005).

Overall, the best fitting structural model assumed that episodic memory predicts strategy selection in the multiplicative judgment task and this choice, in turn, influences judgment accuracy in the multiplicative judgment task, $\chi^2(100) = 94.94$, SF $= 1.03$, $p = .624$, CFI $= 1.00$, RMSEA $= .00$, SRMR $= .05$. This model fit significantly better than a model that did not assume a path from memory abilities to strategy selection or from strategy selection to judgment performance, $\chi^2(102) = 186.60$, SF $= 1.04$, $p < .001$, CFI $= .95$, RMSEA $= .06$, SRMR $= .10$, $\Delta\chi^2(2) = 83.75$, $p < .001$. The model fit could not be improved by additionally assuming that strategy selection mediates the relationship between working memory and judgment accuracy in the linear judgment task, $\chi^2(98) = 90.73$, SF $= 1.03$, $p = .686$, CFI $= 1.00$, RMSEA $= .00$, SRMR $= .04$, $\Delta\chi^2(2) = 4.11$, $p = .128$. Also, assuming that working memory predicts strategy selection in the multiplicative task did not increase model fit, $\chi^2(99) = 93.78$, SF $= 1.03$, $p = .629$, CFI $= 1.00$, RMSEA $= .00$, SRMR $= .04$, $\Delta\chi^2(1) = 1.17$, $p = .280$.

As illustrated in Figure 5, the best fitting structural model shows that strategy selection partly mediated the relationship between episodic memory and judgment performance in the multiplicative task. People with better episodic memory were more likely to select an exemplar-based strategy in the multiplicative task, and this change in judgment strategy increased judgment accuracy in the multiplicative task ($r = -.16$ for the indirect effect, $p < .001$). Better episodic memory still predicted higher judgment accuracy, but the standardized regression weight dropped from $r = -.43$ to $r = -.27$ when the strategy weight in the multiplicative task (called "strategy" in the structural model) was added. In contrast, higher working memory capacity did not

increase the probability of selecting a rule-based strategy in the linear task and strategy selection did not affect judgment performance in the linear task.

Strategy execution. In the Introduction we argued that memory abilities may predict judgment performance because memory abilities improve strategy execution. Specifically, high working-memory capacity may help people execute rule-based strategies, and in turn, strategy execution may mediate the relationship between working memory capacity and judgment accuracy in the linear task. In contrast, episodic memory may help people execute exemplar-based strategies, and in turn, strategy execution may mediate the relationship between episodic memory and judgment accuracy in the multiplicative task. As a further test of these hypotheses, we examined how strategy execution contributes to the relationship between memory skills and judgment performance. In a first step, we determined an indicator for strategy execution in the linear and the multiplicative judgment task based on the computational modeling. To derive this measure, we weighted the predictions of the exemplar and the linear model by the strategy weights (Equation 8) and calculated the RMSD between the weighted predictions and participants' mean judgments. Consequently, the measure determines how consistently people apply the strategy, learned in training, to validation items in the test phase.

To understand how strategy execution is related to memory skills and judgment accuracy, we again relied on mediation analyses. Matching the analysis for strategy selection, we estimated a null model that assumed strategy execution does not mediate the relationship between memory abilities and judgment accuracy. We compared this model to different competitors that assumed (a) strategy execution mediates the relationship between working memory and performance only in the linear judgment task, (b) strategy execution mediates the relationship between episodic memory and performance only in the multiplicative judgment task, or (c) strategy execution

mediates both the relationship between working memory and judgment performance in the linear task and the relationship between episodic memory and judgment performance in the multiplicative task.

The best fitting structural model included a mediation effect of strategy execution on the relationship between working memory and judgment accuracy in the linear judgment task, $\chi^2(100) = 102.57$, SF = 1.05, $p = .410$, CFI = 1.00, RMSEA = .01, SRMR = .04. According to this model, working memory predicts strategy execution in the linear judgment task; hence, the more closely participants followed the strategy learned in training, the more accurate were their judgments. A structural model assuming that strategy execution additionally mediates the relationship between episodic memory and judgment accuracy in the multiplicative judgment task did not improve model fit, $\chi^2(98) = 96.75$, SF = 1.05, $p = .517$, CFI = 1.00, RMSEA = .00, SRMR = .04, $\Delta\chi^2(2) = 5.35$, $p = .069$. Discarding the indirect effect of strategy execution in the linear task, however, significantly harmed the fit of the structural model, $\chi^2(102) = 208.13$, SF = 1.05, $p < .001$, CFI = .94, RMSEA = .06, SRMR = .10, $\Delta\chi^2(2) = 105.56$, $p < .001$.

Figure 6 shows the resulting structural model. In this model, working memory capacity again directly predicts judgment accuracy in the linear task, but to a smaller extent (the standardized regression weight fell from $r = -.35$ to $r = -.24$). Strategy execution mediates this relationship between working memory and judgment accuracy. Higher working memory capacity facilitates executing the learned strategy in linear judgment tasks, and strategy execution, in turn, predicts how accurately people make judgments in linear tasks ($r = -.11$ for the indirect effect, $p = .019$). In the multiplicative task, however, episodic memory does not predict how well people execute a learned strategy, and strategy execution does not lead to more accurate judgments.

General Discussion

Working memory and long-term memory are indispensable for many everyday activities. In fact, working memory capacity predicts performance differences for a wide range of cognitive tasks ranging from reading (Daneman & Carpenter, 1980) to reasoning (Kane et al., 2004; Kyllonen & Cristal, 1990) and also predicts everyday cognitive failures (Unsworth et al., 2012). Likewise, episodic long-term memory has proved useful as an indicator of general intelligence (Jäger et al., 1997). However, little attention has been paid to the question of how various memory abilities influence judgment and decision making (Ashby & O'Brien, 2005; Del Missier et al., 2013; Tomlinson et al., 2011). Our study aimed to fill this gap by investigating how working memory, episodic memory, and implicit memory promote judgment strategies and judgment performance in two kinds of judgment tasks: a linear judgment task that can best be solved by a rule-based cue abstraction strategy and a multiplicative judgment task in which people should rely more often on an exemplar-based strategy. As predicted, we found that working memory capacity was linked to judgment accuracy in linear judgment tasks in which most people tried to abstract rules. In contrast, episodic memory was related to judgment accuracy in multiplicative judgment tasks in which most people relied on exemplar-based strategies. Largely in line with theories in judgment and decision making — and even more with categorization theories (Ashby & O'Brien, 2005; Juslin et al., 2008; Smith et al., 1998) — these results suggest that rule-based and exemplar-based judgment strategies tap into different memory abilities.

The Influence of Memory Abilities on Rule-based Strategies

Rule-based judgment strategies have often been understood as serial, capacity-constrained, hypothesis-testing processes that demand high working memory capacity (Ashby & O'Brien, 2005; Brehmer, 1994; Juslin et al., 2003, 2008). Supporting the idea that working memory capacity is indispensable for making rule-based judgments,

we found that working memory was related to judgment accuracy in linear judgment

tasks in which participants' judgments were, overall, best described by a rule-based

cue abstraction strategy. This result resonates well with previous findings showing that

successfully adopting a rule-based strategy is impeded by cognitive load (Filoteo et

al., 2010; Hoffmann et al., 2013a). Theoretically, two major components of rule-based

judgment strategies contribute to the relationship between working memory capacity

and judgment accuracy. First, abstracting linear rules may require maintaining the

previous judgment object in working memory and comparing it to the current

judgment object (Juslin et al., 2008; Pachur & Olsson, 2012). Second, executing a

rule-based strategy may involve mental updating of the judgment estimate and

inhibiting irrelevant cue information. In line with the latter idea, we found that

working memory capacity promoted executing the chosen strategy more consistently

in linear judgment tasks, and strategy execution, in turn, predicted judgment accuracy.

This finding matches previous research suggesting that working memory capacity

influences how accurately people apply decision rules (Del Missier et al., 2013).

Our results, however, seem to contradict findings by Rolinson, Evans, Walsh,

and Dennis (2011) suggesting that working memory capacity is required only for

learning negative, and not positive relationships between the cues and the criterion. In

contrast, we found that working memory also predicted how successful people were at

learning positive cue–criterion relationships. One explanation could be that our task

was more difficult because the criterion had to be predicted with four instead of only

two cues. Possibly, people with low working memory capacity can still test

hypotheses about two cues, whereas only high working memory capacity allows

people to consider more alternative hypotheses (Dougherty & Hunter, 2003).

Episodic memory, in our study, did not directly predict judgment accuracy in

linear judgment tasks, suggesting that episodic memory is less important than working

memory capacity for making judgments with a cue abstraction strategy. However, memory skills are not independent of each other. Replicating findings from memory research (Del Missier et al., 2013; Mogle et al., 2008; Unsworth, 2010), we found that working memory and episodic memory are moderately correlated, probably reflecting that working memory is needed to encode and retrieve information from long-term memory. Consequently, episodic memory was indirectly related to accuracy in linear judgment tasks through its correlation with working memory ($r = -.14$). Possibly, this indirect relationship suggests that episodic memory is still needed to retrieve cue weights when making a judgment.

One question we did not address is if procedural memory, another type of implicit memory, contributes to rule-based judgment strategies. Procedural memory underlies the learning of motor skills (Squire & Zola, 1996; Willingham, 1998), whereas our measures of implicit memory focused on processing advantages for previously encountered perceptual stimuli. Procedural memory is supposed to underlie the learning of "structured categories containing many exemplars that could not be easily learned via a logical reasoning process" (Ashby & O'Brien, 2005, p. 86). In these information-integration tasks, learning requires many repetitions and the optimal strategy is difficult to verbalize (Ashby & Maddox, 2005). Structurally, information-integration tasks in categorization are most similar to linear, additive judgment tasks. However, it is unlikely that learning to solve additive judgment tasks builds on procedural memory. Not only do people test specific hypothesis when learning to solve these tasks (Brehmer, 1994; Juslin et al., 2008), but they also acquire explicit knowledge about the importance of the cues (Lagnado et al., 2006).

The Influence of Memory Abilities on Exemplar-based Strategies

Surprisingly few studies have empirically investigated how episodic memory is linked to strategies and performance in judgments or decision making. Our study

emphasizes how important episodic memory is for making exemplar-based judgments. We found clear evidence that episodic memory predicts judgment accuracy in multiplicative judgment tasks in which participants' judgments were mostly best described by an exemplar-based judgment strategy. This result is in line with previous studies suggesting that people engage in a strategic memorization process when adopting exemplar-based strategies (Juslin et al., 2008; Olsson et al., 2006) and further supports the theoretical link between episodic memory trace models and exemplar models (Hintzman, 1984, 1986).

In our study we did not find any relationship between implicit memory and exemplar-based judgments, suggesting that implicit memory does not influence judgments. However, it is possible that the lack of a finding was caused by the unreliability of implicit memory measures (Buchner & Brandt, 2003; Buchner & Wippich, 2000; Meier & Perrig, 2000). Although we used several established tasks that should measure implicit memory, correlations between the implicit memory tasks were low and two out of three were not different from zero. This lack of reliability also restricts possible relations to other constructs, making it difficult to interpret our findings. Accordingly, the relation between implicit memory and exemplar-based judgments still remains unclear.

In our study, we found no direct link between working memory capacity and exemplar-based judgments. At first glance, this result seems to contradict previous studies that found working memory helps in solving different kinds of judgment and categorization tasks (Craig & Lewandowsky, 2012; Lewandowsky, 2011; Lewandowsky et al., 2012; Weaver & Stewart, 2012). Indeed, our study differed in some respects from previous research in categorization. While our study investigated how successfully people generalized their performance to new items, previous studies focused mostly on the learning process. In Lewandowsky's (2011) study, for instance,

a learning parameter best captured variations in working memory capacity across six

different categorization tasks. In addition, we assessed judgment performance —

because of time restrictions —with only two different tasks, using judgment accuracy

in the four test blocks as manifest indicators. Accordingly, our measurement focused

more strongly on variance specific to each judgment task, whereas past research

concentrated on the variance shared among different judgment or categorization tasks

(Craig & Lewandowsky, 2012; Lewandowsky, 2011; Lewandowsky et al., 2012;

Weaver & Stewart, 2012). Hence, it is possible that learning in rule- and exemplar-

based judgment tasks requires working-memory capacity, whereas executing a learned

judgment strategy depends on working memory capacity only for rule-based

judgments. However, as mentioned above, working memory capacity was moderately

correlated with episodic memory in our study. Accordingly, working memory was

helpful not only for solving linear judgment tasks, but also for solving multiplicative

judgment tasks: Higher working memory capacity predicted higher judgment accuracy

in the multiplicative judgment task through its connection to episodic memory ($r = -$

$.17$). Apparently, successfully solving judgment tasks relies on the interplay between

episodic memory and working memory — an interpretation that is generally in line

with the idea that learning in a huge variety of judgment tasks depends on working

memory capacity (Weaver & Stewart, 2012).

Memory Abilities and Strategy Use

In the past decade, judgment research has focused mostly on task

characteristics as a determinant of judgment strategies (Juslin et al., 2003, 2008;

Karlsson et al., 2007; von Helversen et al., 2013; von Helversen & Rieskamp, 2009).

Consistent with prior research we found that most participants relied on a rule-based

cue abstraction strategy in a linear judgment task and shifted to exemplar-based

strategies in multiplicative judgment tasks (Hoffmann et al., 2013a; Juslin et al., 2008;

Karlsson et al., 2007). However, individual differences, such as age or intelligence,

can also drive shifts between different types of strategies (Bröder, 2003; Mata, von

Helversen et al., 2012). Specifically, we argued that memory abilities may influence

not only how well people execute a strategy but also which strategies people select

(Beach & Mitchell, 1978; Lemaire & Siegler, 1995; Mata, Pachur et al., 2012).

Whereas neither working memory capacity nor episodic memory influenced strategy

selection in the linear task, episodic memory fostered the probability of selecting an

exemplar strategy in the multiplicative task. Furthermore, strategy selection partly

mediated the relationship between episodic memory and judgment performance. This

result dovetails with the idea that memory abilities may reduce the costs associated

with a strategy and, in turn, increase the preference for employing a specific strategy

(Beach & Mitchell, 1978; Payne et al., 1993; Rieskamp & Otto, 2004).

Following the strategy selection approach, however, one would have expected

that working memory capacity predicts as well to what extent people select a rule-

based strategy in the linear task. One reason why we did not find this relationship

could be that rule-based strategies act as a default (Karlsson et al., 2008; Olsson et al.,

2006). In line with this argumentation, few people chose an exemplar strategy in the

linear tasks. Consequently, only engaging in exemplar-based memorization processes

would require an active choice, whereas the success of rule-based strategies may

depend more on the effort needed to execute the strategy. This explanation is

supported by the finding that working memory capacity predicted how well the

learned strategy was executed in the linear task, suggesting that the inability to

accurately use a strategy does not necessarily lead to a strategy shift. In contrast, how

well the learned strategy was executed in the multiplicative task was unrelated to

episodic memory, suggesting that those participants who did not shift to the task-

appropriate exemplar-based strategy nevertheless applied the rules they learned consistently.

In sum, our results demonstrate that episodic memory plays an important role in strategy selection (Mata, von Helversen et al., 2012) but do not provide any evidence that working memory capacity — as previously suggested — predicts more adaptive strategy selections (Bröder, 2003; Craig & Lewandowsky, 2012; Lewandowsky et al., 2012; Mata, Pachur et al., 2012). These results emphasize that reducing strategy selection to a question of working memory capacity probably oversimplifies the idea of adaptive strategy use. Current research, for instance, proposes that people have stable preferences across domains for learning based on rules or on exemplars (McDaniel, Cahill, Robbins, & Wiener, 2013). Investigating these preferences in conjunction with cognitive abilities hopefully allows researchers to form a more comprehensive picture of how task demands and characteristics of the decision maker constrain the repertoire of applicable strategies.

Conclusions

Twenty years ago, Elke Weber and colleagues (1995) reminded us that we should not forget memory processes when thinking about how people make judgments. Our results suggest that different judgment strategies take advantage of specific memory processes: Whereas rule-based strategies draw on working memory capacity, exemplar-based strategies exploit encoding and retrieval processes in episodic long-term memory. Thus, knowledge about working memory and long-term memory processes may help explain how people successfully solve judgment tasks that range from simple daily judgments such as estimating the probability of rainfall to professional judgments such as judging the quality of a job candidate.

Footnotes

1. In a pilot study, 12 participants rated 100 German sentences for plausibility. Only highly plausible or implausible sentences were included in the final reading span test.

2. In a pilot study, we included a threshold procedure using 40 independent nouns. The results showed that participants correctly identified half of the nouns using a mask with nine black squares so that 56% of the noun was masked.

3. To assure that old and new sounds were equally easy to identify among distractors, we conducted a pilot study with 24 subjects. In this pilot study, half of the participants heard half of the sounds without noise in the study phase; the other half of the participants heard the remaining sounds in the study phase. Afterward, old and new sounds were presented embedded in noise and participants were asked to identify them among two distractors. For the final experiment, old and new sounds were matched on performance for old sounds.

4. We also fitted an exemplar model with four attention parameters to participants' judgments. However, replicating results from previous studies (Hoffmann et al., 2013a; von Helversen & Rieskamp, 2008), this model failed to outperform an exemplar model with one parameter in predicting participants' judgments for validation items in either the linear task (RMSD = 5.3) or the multiplicative task (RMSD = 5.85).

5. Judgment accuracy was measured in RMSD with lower RMSD indicating more accurate judgments. Accordingly, negative correlations imply that higher working memory predicts higher judgment accuracy.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in

practice: A review and recommended two-step approach. *Psychological*

Bulletin, 103, 411–423. doi: 10.1037/0033-2909.103.3.411

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review*

of Psychology, 56, 149–178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory

systems. *Trends in Cognitive Sciences, 9*, 83–89.

doi:10.1016/j.tics.2004.12.003

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 (Vol. 2, pp. 89–195). New York, NY: Academic

Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.),

The psychology of learning & motivation (Vol. 8, pp. 47–89). New York,

NY: Academic Press.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and

retrieval from long-term memory. *Journal of Experimental Psychology:*

General, 113, 518–540. doi: 10.1037//0096-3445.113.4.518

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of

decision strategies. *The Academy of Management Review*, *3*, 439–449.

Berry, C. J., Shanks, D. R., Speekenbrink, M., & Henson, R. N. A. (2012). Models

of recognition, repetition priming, and fluency: Exploring a new framework.

Psychological Review, 119, 40–79. doi: 10.1037/a0025464

Brehmer, B. (1994). The psychology of linear judgement models. *Acta*

Psychologica, 87, 137–154. doi: 10.1016/0001-6918(94)90048-5

Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language, 66*, 407–415. doi: 10.1016/j.jml.2011.12.009

Brickenkamp, R. (2002). *Test d2. Aufmerksamkeits-Belastungs-Test*. Göttingen, Germany: Hogrefe.

Bröder, A. (2003). Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 611–625. doi:10.1037/0278-7393.29.4.611

Brown, S., & Craik, F. (2000). Encoding and retrieval of information. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 93–107). New York, NY: Oxford University Press.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York, NY: Wiley.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press.

Bruss, P. J., & Mitchell, D. B. (2009). Memory systems, processes, and tasks: Taxonomic clarification via factor analysis. *The American Journal of Psychology*, *122*, 175–189.

Buchner, A., & Brandt, M. (2003). Further evidence for systematic reliability differences between explicit and implicit memory tests. *The Quarterly Journal of Experimental Psychology A*, *56*, 193–209. doi: 10.1080/02724980244000260

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40*, 227–259. doi: 10.1006/cogp.1999.0731

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44,* 171–189. doi: 10.1006/jmps.1999.1282

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. doi: 10.1037/0033-2909.132.3.354

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786. doi: 10.3758/BF03196772

Cooksey, R. W. (1996). The methodology of social judgement theory. *Thinking & Reasoning, 2*, 141–174. doi: 10.1080/135467896394483

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *The Quarterly Journal of Experimental Psychology, 65,* 439–464. doi: 10.1080/17470218.2011.608854

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*, 159–180. doi: 10.1037/0096-3445.125.2.159

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466. doi: 10.1016/S0022-5371(80)90312-6d

Davis, T., Love, B. C., & Preston, A. R. (2012). Learning the exception to the rule: Model-based FMRI reveals specialized representations for surprising category members. *Cerebral Cortex, 22*, 260–273. doi:10.1093/cercor/bhr036

Del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2010). Executive functions in

decision making: An individual differences approach. *Thinking & Reasoning,*

16, 69–97. doi: 10.1080/13546781003630117

Del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2012). Decision-making

competence, executive functioning, and general cognitive abilities. *Journal of*

Behavioral Decision Making, 25, 331–351. doi:1 0.1002/bdm.731

Del Missier, F., Mäntylä, T., Hansson, P., Bruine de Bruin, W., Parker, A. M., &

Nilsson, L.-G. (2013). The multifold relationship between memory and

decision making: An individual-differences study. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 39, 1344–1364. doi:

10.1037/a0032379

Dew, I. T. Z., & Cabeza, R. (2011). The porous boundaries between explicit and

implicit memory: Behavioral and neural evidence. *Annals of the New York*

Academy of Sciences, *1224*, 174–190. doi: 10.1111/j.1749-6632.2010.05946.x

Dougherty, M. R. P., & Hunter, J. E. (2003). Hypothesis generation, probability

judgment, and individual differences in working memory capacity. *Acta*

Psychologica, *113*, 263–282. doi: 10.1016/S0001-6918(03)00033-7

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and

process-tracing models of judgment. *Psychological Review, 86*, 465–485.

doi: 10.1037//0033-295X.86.5.465

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of*

factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category

learning. *Journal of Experimental Psychology: General, 127*, 107–140. doi:

10.1037/0096-3445.127.2.107

Estes, W. K. (1986). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115*, 155–174.

Filoteo, J. V., Lauritzen, J. S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science, 21*, 415–423. doi: 10.1177/0956797610362646

Fleischman, D. A., Wilson, R. S., Gabrieli, J. D. E., Bienias, J. L., & Bennett, D. A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Aging, 19*, 617–625. doi: 10.1037/0882-7974.19.4.617

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 164–178. doi: 10.1037/0278-7393.10.1.164

Graves, L., & Karren, R. (1992). Interviewer decision processes and effectiveness: An experimental policy-capturing investigation. *Personnel Psychology, 45*, 313–340. doi: 10.1111/j.1744-6570.1992.tb00852.x

Gygi, B., & Shafiro, V. (2010). Development of the database for environmental sound research and application (DESRA): Design, functionality, and retrieval considerations. *EURASIP Journal on Audio, Speech, and Music Processing, 2010*, 1–12. doi: 10.1155/2010/654914

Hager, W., & Hasselhorn, M. (Eds.). (1994). *Handbuch deutschsprachiger Wortnormen*. Göttingen: Hogrefe.

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition, 65*, 197–230. doi: 10.1016/S0010-0277(97)00044-9

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! Testing! One, two, three — Testing the theory in structural equation models! *Personality and Individual Differences, 42*, 841–850. doi: 10.1016/j.paid.2006.10.001

Helsdingen, A. S., van Gog, T., & van Merriënboer, J. J. G. (2011). The effects of practice schedule on learning a complex judgment task. *Learning and Instruction*, *21*, 126–136. doi: 10.1016/j.learninstruc.2009.12.001

Hintzman, D. L. (1984). MINERVA 2 : A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*, 96–101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428. doi: 10.1037//0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551. doi:10.1037//0033-295X.95.4.528

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013a). Deliberation's blindsight: How cognitive load can improve judgments. *Psychological Science, 26,* 869–879. doi: 10.1177/0956797612463581

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013b). *From rules to exemplars: Similar task features shape judgment and categorization processes*. Manuscript in preparation.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*, 1–55.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*, 90–98. doi: 10.1016/j.jcps.2009.09.003

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. Form 4*. Göttingen, Germany: Hogrefe.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*, 259–298. doi: 10.1016/j.cognition.2007.02.003

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132*, 133–156. doi: 10.1037/0096-3445.132.1.133

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217. doi:10.1037/0096-3445.133.2.189

Karlsson, L., Juslin, P., & Olsson, H. (2007). Adaptive changes between cue abstraction and exemplar memory in a multiple-cue judgment task with continuous cues. *Psychonomic Bulletin & Review, 14*, 1140–1146.

Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making : Contingent, not automatic, strategy shifts? *Judgment and Decision Making, 3*, 244–260.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford.

Knowlton, B. J. (1999). What can neuropsychology tell us about category learning? *Trends in Cognitive Sciences*, *3*, 123–124. doi: 10.1016/S1364-6613(99)01292-9

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262,* 1747–1749. doi: 10.1126/science.8259522

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*, 389–433. doi: 10.1016/S0160-2896(05)80012-1

Lagnado, D. A., Newell, B., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple cue learning. *Journal of Experimental Psychology: General, 135*, 162–183. doi: 10.1037/0096-3445.135.2.162

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124*, 83–97. doi: 10.1037//0096-3445.124.1.83

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 720–738. doi: 10.1037/a0022639

Lewandowsky, S., Yang, L. X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*, 881–904. doi: 10.1037/a0027298

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review, 118*, 393–437. doi:10.1037/a0024143

Martin, E. (1968). Stimulus meaningfulness and paired-associates transfer: An encoding variability hypothesis. *Psychological Review, 75*, 421–441.

Mata, R., Pachur, T., von Helversen, B., Hertwig, R., Rieskamp, J., & Schooler, L. (2012). Ecological rationality: A framework for understanding and aiding the

aging decision maker. *Frontiers in Neuroscience, 6*, 1–19. doi: 10.3389/fnins.2012.00019

Mata, R., von Helversen, B., Karlsson, L., & Cüpper, L. (2012). Adult age differences in categorization and multiple-cue judgment. *Developmental Psychology*, *48*, 1188–1201. doi: 10.1037/a0026084

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2013). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General.* Advance online publication. doi:10.1037/a0032963

McDaniel, M. A., Fadler, C. L., & Pashler, H. (2013). Effects of spaced versus massed training in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. doi: 10.1037/a0032184

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238. doi:10.1037/0033-295X.85.3.207

Meier, B., & Perrig, W. J. (2000). Low reliability of perceptual priming: Consequences for the interpretation of functional dissociations between explicit and implicit memory. *Quarterly Journal of Experimental Psychology: A Human Experimental Psychology, 53*, 211–233. doi: 10.1080/713755878

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100. doi: 10.1006/cogp.1999.0734

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? An examination of the relationships among

working memory, secondary memory, and fluid intelligence. *Psychological Science, 19*, 1071–1077. doi: 10.1111/j.1467-9280.2008.02202.x

Musen, G., & Treisman, A. (1990). Implicit and explicit memory for visual patterns. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *16*, 127–137. doi: 10.1037/0278-7393.16.1.127

Newell, B., Dunn, J., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 167–215). Burlington, MA: Academic Press.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 700–708. doi: 10.1037/0278-7393.14.4.700

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53–79. doi:10.1037/0033-295X.101.1.53

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*, 247–255. doi: 10.1111/1467-9280.00051

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, pp. 45–100). Burlington, MA: Academic Press.

Olsson, A.-C., Enkvist, T., & Juslin, P. (2006). Go with the flow: How to master a nonlinear multiple-cue judgment task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1371–1384. doi: 10.1037/0278-7393.32.6.1371

Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology, 65*, 207–240. doi:10.1016/j.cogpsych.2012.03.003

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 548–568. doi: 10.1037/0278-7393.21.3.548

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.

Perruchet, P., & Beaveux, P. (1989). Correlational analyses of explicit and implicit memory performance. *Memory & Cognition, 17*, 77–86. doi: 10.3758/BF03199559

Platzer, C., & Bröder, A. (2013). When the rule is ruled out: Exemplars and rules in decisions from memory. *Journal of Behavioral Decision Making*, *26*, 429–441. doi: 10.1002/bdm.1776

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*, 164–171. doi: 10.1027/1015-5759/a000123

Rieskamp, J. (2006). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1355–1370. doi:10.1037/0278-7393.32.6.1355

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*, 207–236. doi: 10.1037/0096-3445.135.2.207

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*, 1043–1056. doi: 10.1037/0003-066X.45.9.1043

Rolison, J. J., Evans, J. S. B. T., Walsh, C. R., & Dennis, I. (2011). The role of working memory capacity in multiple-cue probability learning. *Quarterly Journal of Experimental Psychology*, *64*, 1494–1514. doi: 10.1080/17470218.2011.559586

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*, 211–227. doi: 10.1037/0096-3445.126.3.211

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217–236.

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, *133*, 63–82. doi: 10.1037/0096-3445.133.1.63

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General, 133*, 534–553. doi:10.1037/0096-3445.133.4.534

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514.

Shiffrin, R., & Atkinson, R. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76*, 179–193.

Smith, E. E. (2008). The case for implicit category learning. *Cognitive, Affective, & Behavioral Neuroscience, 8*, 3–16. doi:10.3758/CABN.8.1.3

Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience and Biobehavioral Reviews, 32*, 249–264. doi:10.1016/j.neubiorev.2007.07.009

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65*, 167–196. doi: 10.1016/S0010-0277(97)00043-7

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences of the United States of America, 93*, 13515–13522.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1*, 31–65. doi: 10.1146/annurev.clinpsy.1.102803.144239

Tomlinson, T., Marewski, J. N., & Dougherty, M. (2011). Four challenges for cognitive research on the recognition heuristic and a call for a research strategy shift. *Judgment and Decision Making, 6*, 89–99.

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica, 134*, 16–28. doi: 10.1016/ j.actpsy.2009.11.010

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41*, 242–254. doi: 10.3758/s13421-012-0261-x

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104–132. doi: 10.1037/0033-295X.114.1.104

Unsworth, N., McMillan, B. D., Brewer, G. A., & Spillers, G. J. (2012). Everyday

    attention failures: An individual differences investigation. *Journal of*

    Experimental Psychology: Learning, Memory, and Cognition, 38, 1765–

    1772. doi: 10.1037/a0028075

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W.

    (2009). Complex working memory span tasks and higher-order cognition: A

    latent-variable analysis of the relationship between processing and storage.

    Memory, 17, 635–654. doi: 10.1080/09658210902998047

von Helversen, B., Karlsson, L., Mata, R., & Wilke, A. (2013). Why does cue polarity

    information provide benefits in inference problems? The role of strategy

    selection and knowledge of cue importance. *Acta Psychologica, 144*, 73–82.

    doi:10.1016/j.actpsy.2013.05.007

von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking

    beyond looks? From similarity-based to cue abstraction processes in

    multiple-cue judgment. *Developmental Psychology, 46*, 220–229.

    doi:10.1037/a0016690

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory

    of quantitative estimation. *Journal of Experimental Psychology: General*,

    137, 73–96. doi: 10.1037/0096-3445.137.1.73

von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-

    based and exemplar-based processes compared. *Journal of Experimental*

    Psychology: Learning, Memory, and Cognition, 35, 867–889.

    doi:10.1037/a0015501

Weaver, E. A., & Stewart, T. R. (2012). Dimensions of judgment: Factor analysis of

    individual differences. *Journal of Behavioral Decision Making, 25*, 402–413.

    doi: 10.1002/bdm.748

Weber, E. U., Goldstein, W. M., & Barlas, S. (1995). And let us not forget memory: The role of memory processes and techniques in the study of judgment and choice. *The Psychology of Learning and Motivation, 32*, 33–81. doi: 10.1016/S0079-7421(08)60307-2

Wigton, R. S. (1996). Social judgement theory and medical judgement. *Thinking and Reasoning, 2,* 175–190. doi: 10.1080/135467896394492

Willingham, D. (1998). A neuropsychological theory of motor skill learning. *Psychological Review, 105*, 558–584. doi: 10.1037/0033-295X.105.3.558

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*, 387–398. doi: 10.3758/BF03193416

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition, 35*, 1380–1398. doi: 10.3758/BF03193609

Table 1

Training and Validation Items Used in the Multiplicative and the Linear Judgment Task. The Judgment Criterion Was Derived from Equation 1 (Linear) and Equation 2 (Multiplicative).

Cue values				Judgment criterion		Set
Cue 1	Cue 2	Cue 3	Cue 4	Multiplicative	Linear	
2	1	0	3	2	14	Training
1	4	1	4	5	22	Training
0	3	1	2	2	13	Training
0	2	3	0	1	12	Training
5	5	4	0	29	43	Training
0	4	5	4	12	26	Training
2	4	3	0	9	26	Training
1	4	3	5	13	27	Training
1	0	2	4	1	12	Training
1	0	0	2	1	6	Training
5	3	3	5	21	40	Training
1	1	5	5	7	22	Training
1	2	0	5	2	15	Training
5	5	0	1	4	36	Training
0	4	3	1	4	19	Training
4	2	1	3	6	27	Training
0	5	2	3	6	22	Training
5	5	2	4	22	43	Training
5	1	3	4	9	33	Training

Cue values				Judgment criterion		Set
Cue 1	Cue 2	Cue 3	Cue 4	Multiplicative	Linear	
4	0	2	4	3	24	Training
1	4	1	5	6	23	Training
3	0	5	5	3	27	Training
0	2	5	0	2	16	Training
1	5	2	4	10	27	Training
3	4	5	5	30	39	Training
3	5	1	4	10	33	Validation
3	4	4	3	21	35	Validation
5	0	3	4	4	30	Validation
3	4	2	5	14	33	Validation
5	0	5	5	4	35	Validation
3	2	0	2	2	20	Validation
2	3	4	0	9	25	Validation
4	5	4	5	36	44	Validation
5	0	5	3	4	33	Validation
4	3	0	1	3	26	Validation
2	1	2	0	3	15	Validation
2	5	2	3	12	30	Validation
4	0	0	2	2	18	Validation
4	1	1	1	4	22	Validation
3	3	3	5	15	32	Validation

Table 2

Descriptive Statistics for the Memory and the Judgment Tasks

Task	M	SD	Skewness	Kurtosis
Operation span	58.4	11.7	-1.3	2.2
Reading span	57.6	11.8	-1.2	2.1
Symmetry span	29.9	7.4	-0.7	0.1
Recognition (% recalled)	86.5	8.8	-0.7	0.2
Cued recall (% recalled)	41.4	19.6	0.3	-0.2
Free recall (% recalled)	44.6	16.5	0.3	-0.1
Speeded presentation (ms)	55	1023	0.2	5.3
Degraded presentation (ms)	-1293	3471	0.4	3.5
Identification in noise (ms)	-371	788	-0.9	3.1
Linear judgment				
Last training block	6.1	2.4	1.4	4.1
Test (Mean)	5.4	1.9	0.5	0.6
Multiplicative judgment				
Last training block	5.3	1.8	0.6	0.4
Test (Mean)	5.1	1.9	0.9	0.5

Table 3

Model Fits in the Linear and the Multiplicative Judgment Task (Standard Deviations in Parentheses)

Model fit	Baseline model	Linear model	Exemplar model
Linear task			
Training RMSD	9.0 (1.3)	4.5 (1.4)	5.3 (1.6)
Test RMSD	7.3 (1.6)	3.8 (1.4)	5.2 (1.5)
Classification (N)	17	220	42
Multiplicative task			
Training RMSD	7.3 (1.2)	4.7 (0.9)	4.7 (1.1)
Test RMSD	6.9 (1.9)	4.6 (1.3)	4.2 (1.1)
Classification (N)	4	99	176

Note. RMSD: root mean square deviation

Table 4

Zero-order Correlations Between All Memory Tasks and Test Performance in the Judgment Tasks

Task	Episodic memory			Working memory			Implicit memory			Additive task				Multiplicative task			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Episodic memory																	
1 Recognition	1																
2 Cued recall	.228	1															
3 Free recall	.205	.126	1														
Working memory																	
4 Operation	.027	.123	.074	1													
5 Reading	.138	.158	.108	.540	1												
6 Symmetry	.168	.145	.111	.398	.306	1											
Implicit memory																	
7 Speeded	-.023	.101	.113	-.045	-.065	.000	1										
8 Degraded	-.033	-.012	-.133	-.043	-.076	.036	.018	1									
9 Noise	-.038	-.031	.067	.073	.021	-.023	.152	-.011	1								

Task	Episodic memory			Working memory			Implicit memory			Additive task				Multiplicative task			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Additive task																	
10 Test Block 1	-.032	-.156	-.076	-.193	-.152	-.152	-.001	.063	-.031	1							
11 Test Block 2	-.040	-.194	-.130	-.239	-.230	-.197	-.037	.053	-.044	.734	1						
12 Test Block 3	-.062	-.202	-.112	-.213	-.214	-.185	-.066	.100	-.038	.738	.799	1					
13 Test Block 4	-.055	-.159	-.062	-.227	-.210	-.184	-.015	.047	-.079	.710	.749	.777	1				
Multiplicative task																	
14 Test Block 1	-.157	-.082	-.202	-.082	-.128	-.058	-.060	.147	-.040	.114	.174	.157	.101	1			
15 Test Block 2	-.174	-.169	-.175	-.148	-.171	-.067	-.061	.070	-.034	.229	.205	.234	.179	.675	1		
16 Test Block 3	-.168	-.100	-.149	-.147	-.174	-.070	-.077	.090	.001	.178	.241	.237	.139	.666	.705	1	
17 Test Block 4	-.136	-.117	-.180	-.100	-.120	-.067	-.022	.123	-.001	.177	.149	.180	.107	.635	.734	.717	1

Figure 1. Strategy classification of participants in the linear and the multiplicative judgment task.

Linear task   Multiplicative task

15%

6%   35%

79%   1%   63%

Linear model
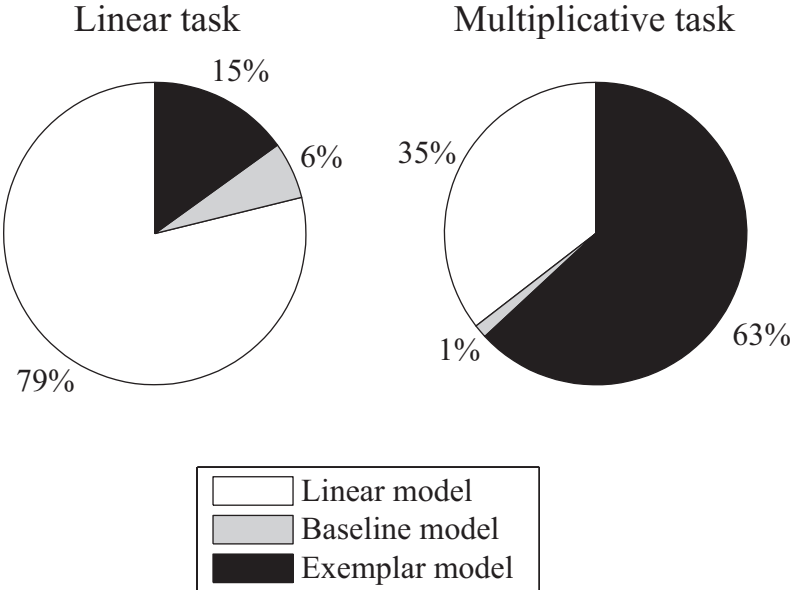Baseline model
Exemplar model

Figure 2. Measurement model for memory abilities with a correlation between the latent constructs working memory and episodic memory. Circles represent latent constructs and squares represent manifest variables. The numbers above the long, single-headed arrows give the standardized factor loadings; the numbers next to the short, single-headed arrows are error variances of the manifest variables. These error variances cover all task-specific variances, including measurement error, material-specific variance, and test-specific variance. Double-headed arrows indicate correlations between the latent constructs. All loadings and correlations are standardized.
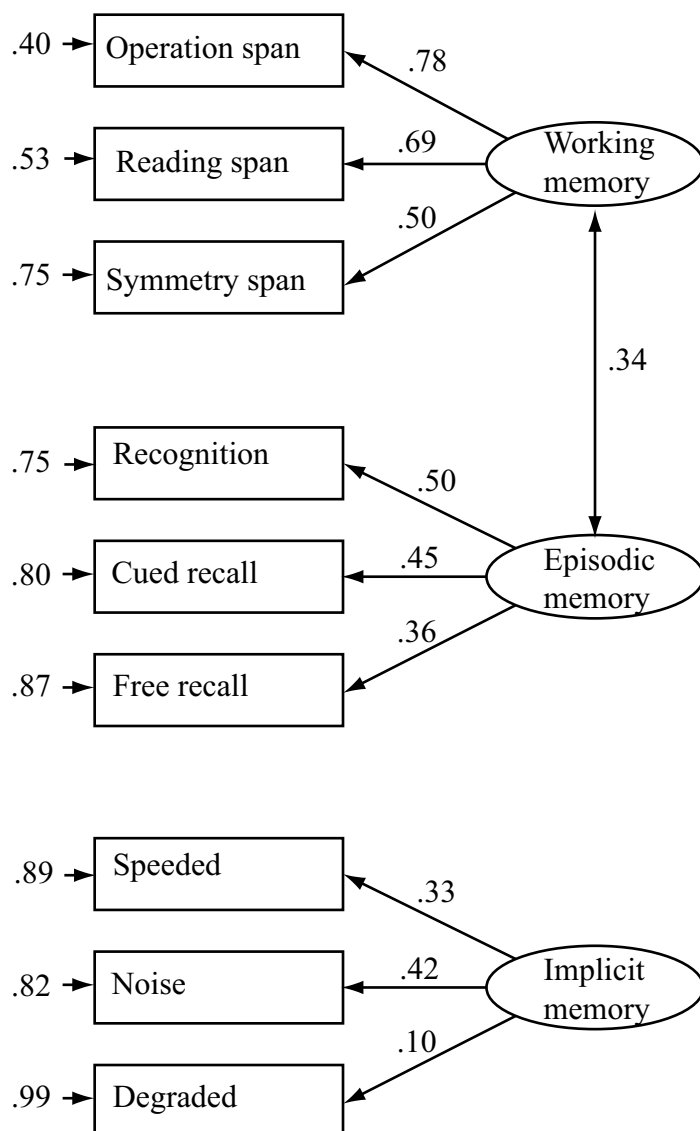
Figure 3. Measurement model for judgment performance for validation items with a correlation between the latent constructs judgment performance in the multiplicative task and judgment performance in the additive task (see Figure 2 for a detailed description of the graphical representation). All loadings and correlations are standardized.
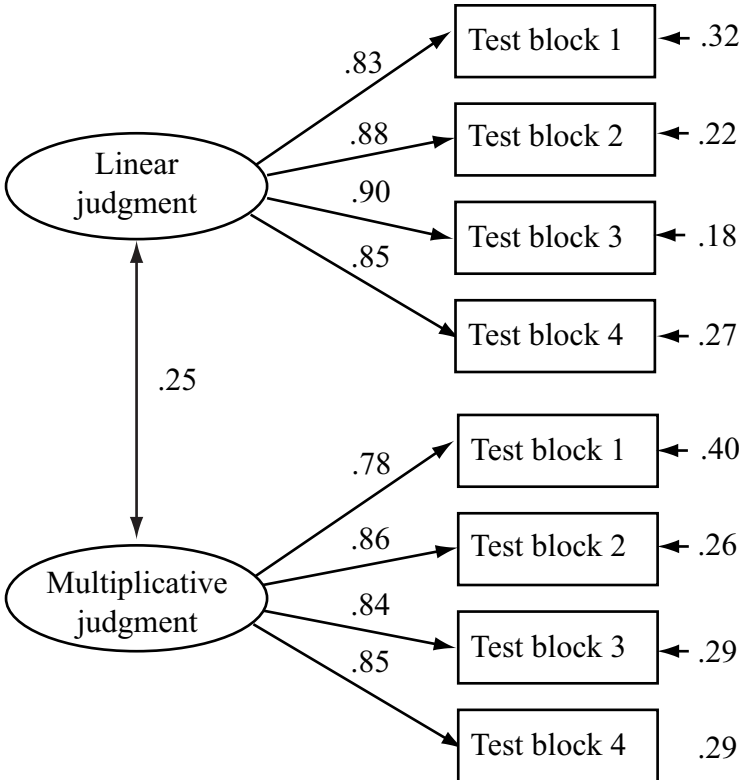
Figure 4. Structural model relating judgment performance in the test phase to memory abilities (see Figure 2 for a detailed description of the graphical representation). Judgment accuracy was measured in root mean square deviation (RMSD) with lower RMSD indicating more accurate judgments. Accordingly, correlations between the memory constructs and judgment accuracy are negative. All loadings and correlations are standardized.
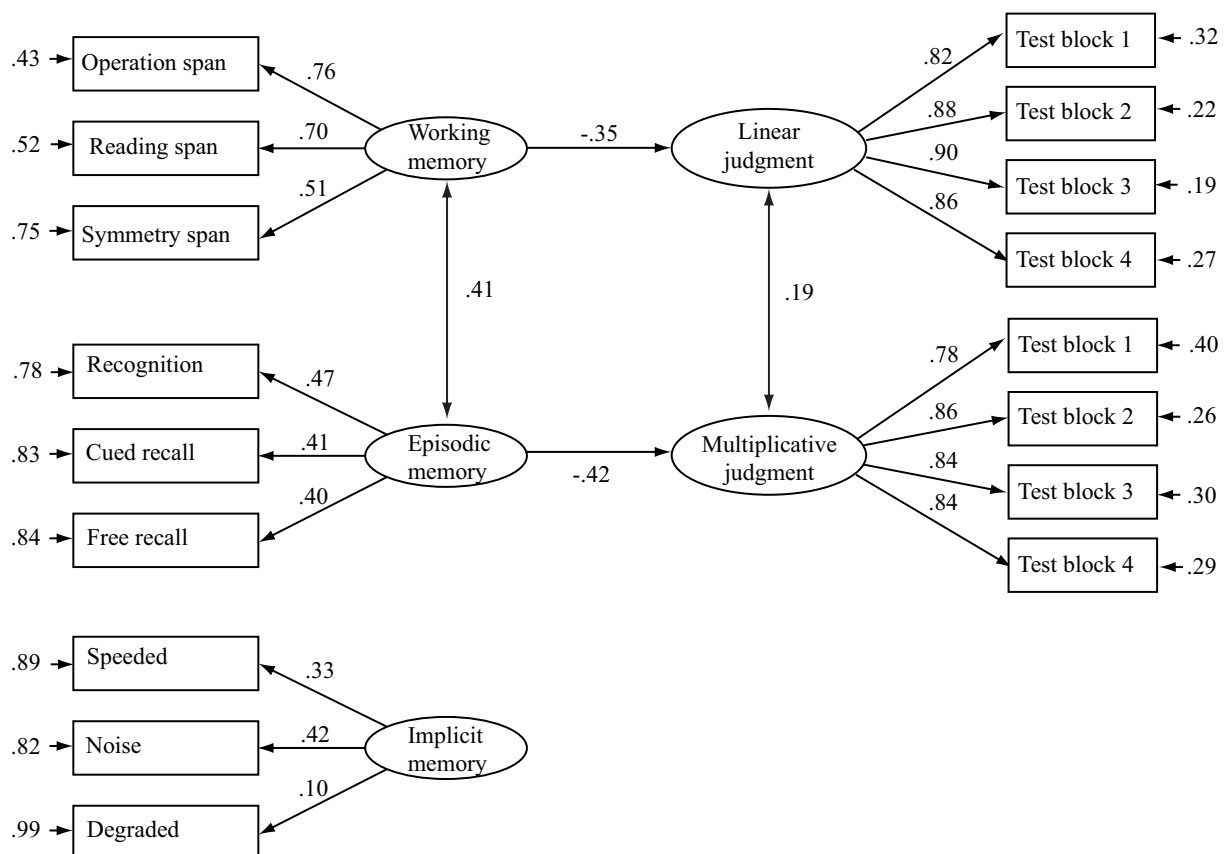
Figure 5. Structural model relating judgment performance in the test phase through strategy selection to memory abilities (see Figure 2 for a detailed description of the graphical representation). All loadings and correlations are standardized. Correlation in parentheses indicates correlation without indirect effect.
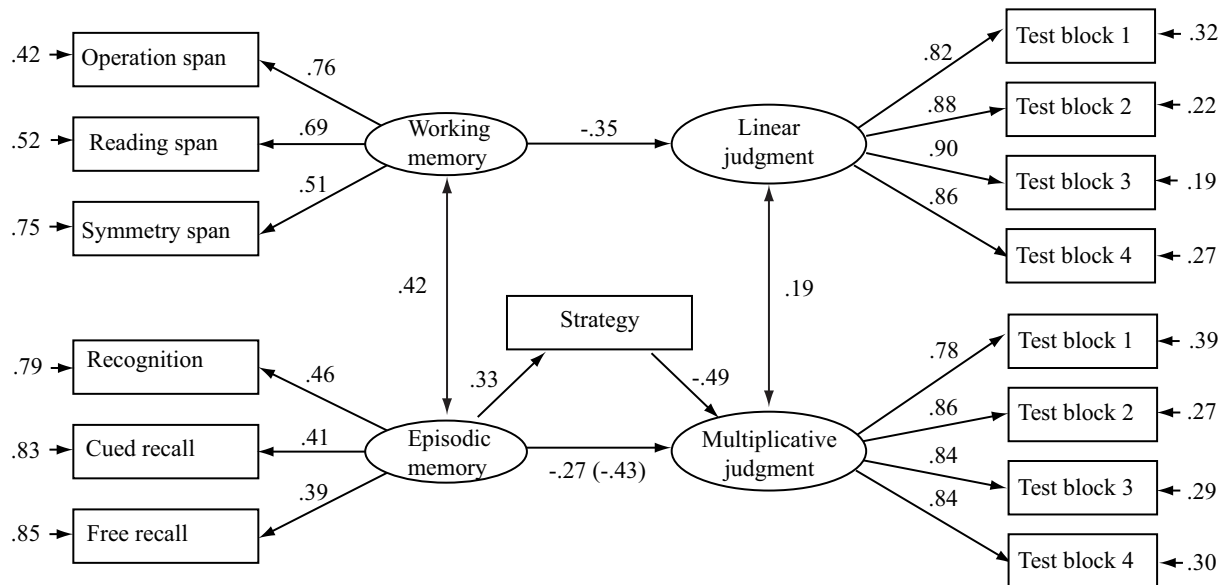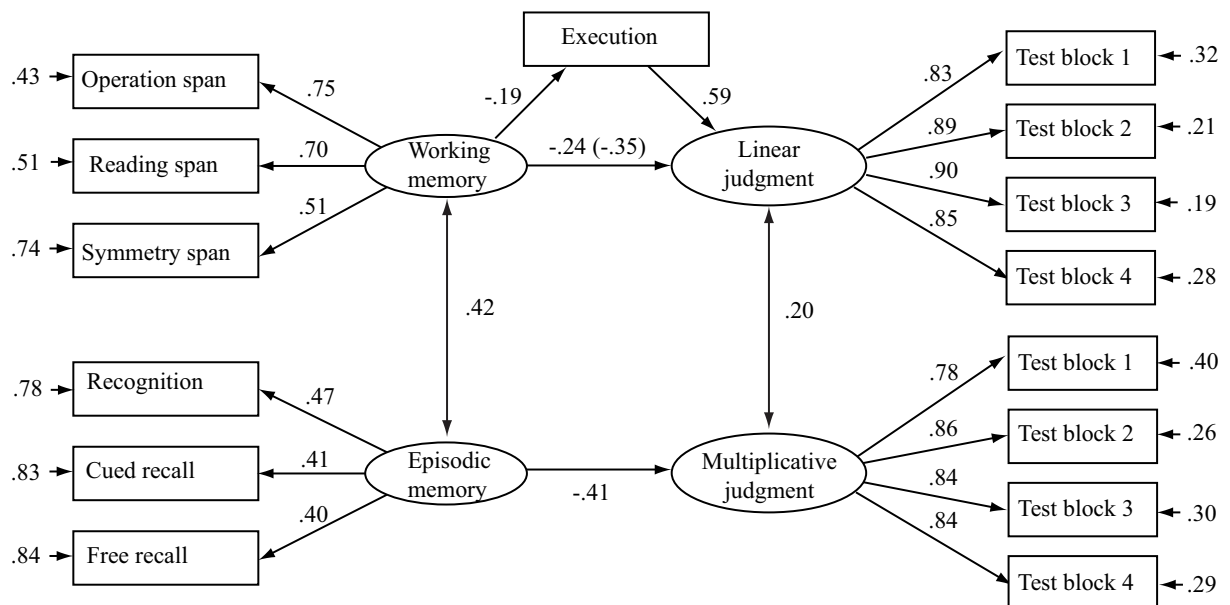
Figure 6. Structural model relating judgment performance in the test phase through strategy consistency to memory abilities (see Figure 2 for a detailed description of the graphical representation). All loadings and correlations are standardized. Correlation in parentheses indicates correlation without indirect effect.

Curriculum Vitae

CONTACT
Janina Hoffmann
Missionsstrasse 62a
4055 Basel
Switzerland

email: janina.hoffmann@unibas.ch

PERSONAL DATA
Date of birth: June 16, 1986
Birth place: Böblingen, Germany

EDUCATION

2014/05	**Dr. phil, summa cum laude**, University of Basel, Switzerland.
2010/08	**Diploma in Psychology**, University of Mannheim, Germany

EMPLOYMENT (SINCE 2004)

2013/08 – 2016/08	Phd student/PostDoc within the project "Modeling human judgment: Integrating memory and rule-based processes"
	Principal investigator: Bettina von Helversen, Jörg Rieskamp
2010/08 - 2013/08	Phd student within the project „The memory foundation of judgment and categorization processes"
	Principal investigator: Bettina von Helversen, Jörg Rieskamp

PUBLICATIONS

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation's blind sight: How cognitive load can improve judgments. *Psychological Science, 24,* 869-879

Hoffmann, J A., von Helversen, B., & Rieskamp, J. (2013). How episodic and working memory affect rule- and memory-based judgments. In Knauff, M., Pauen, M., Sebanz, N., & Wachsmuth, I. (Eds.), *Proceedings of the 35th Annual Con-*

ference of the Cognitive Science Society (pp. 585-590). Austin, TX: Cognitive Science Society.

MANUSCRIPTS

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect judgment performance and strategy use in rule-based and exemplar-based judgment tasks. Manuscript under revision.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Similar task features shape judgment and categorization processes. Manuscript submitted for publication.

CONFERENCE PRESENTATIONS

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Learning rules in judgment: A comparison of the delta-learning rule and linear regression models. Talk presented at the 56. Tagung experimentell arbeitender Psychologen, Giessen, Germany.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). How episodic and working memory affect rule- and exemplar-based judgments. Talk presented at SPUDM 24, Barcelona, Spain.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Learning rules in judgment: A comparison of the delta-learning rule and linear regression models. Poster presented at 46th Annual Meeting of the Society of Mathematical Psychology, Potsdam, Germany.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). How episodic and working memory affect rule- and exemplar-based judgments. Talk presented at the Twelfth Annual Summer Interdisciplinary Conference, Cortina d'Ampezzo, Italy.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). The contribution of memory abilities to rule-based and exemplar-based judgments. Talk presented at the 55. Tagung experimentell arbeitender Psychologen, Vienna, Austria.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). Task complexity encourages shifts between rule- and memory-based strategies in categorization and judgment. Talk presented at the 2012 Meeting of the Society for Judgment and Decision Making, Minneapolis, US.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). Task complexity encourages shifting from rules to exemplars in categorization and judgment. Poster presented at the 53. Meeting of the Psychonomic Society, Minneapolis, US.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). Wenn zuviel Nachdenken hinderlich ist: Wie kognitive Belastung zu besseren Urteilstrategien führen

kann. Talk presented at the 48. Kongress der Deutschen Gesellschaft für Psychologie, Bielefeld, Germany.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). Task complexity encourages shifts between rules and exemplars in categorization and judgment. Talk presented at the 5th Workshop of Young Researchers in Judgment and Decision Making, Basel, Switzerland.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). Learning to categorize: From easy to complex tasks. Talk presented at the 2012 SNF Summer School in Computational Modeling of Cognition, Bergün, Switzerland.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). Complexity encourages processing shifts in categorization. Poster presented at the 2012 SNF Summer School in Computational Modeling of Cognition, Bergün, Switzerland.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). The impact of task complexity on rule- and exemplar-based processes in categorization and judgment. In A. Bröder, E. Erdfelder, B. E. Hilbig, T. Meiser, R. F. Pohl, & D. Stahlberg (Eds.). *Abstracts of the 54. Tagung experimentell arbeitender Psychologen* (p. 290). Lengerich: Pabst.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2011). How distraction can improve judgment processes. Poster presented at the 2011 Meeting of the Society for Judgment and Decision Making, Seattle, US.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2011). How cognitive load can improve judgment processes. Poster presented at the 52. Meeting of the Psychonomic Society, Seattle, US.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2011). Deliberation's blind sight: How does distraction affect judgment processes? Talk presented at SPUDM 23, London Kingston, UK.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2011). How cognitive load improves judgment processes. Talk presented at the 4th Workshop of Young Researchers in Judgment and Decision Making, Bonn, Germany.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2011). Kontrollierte vs. automatische Schätzprozesse: Zur Rolle des Arbeitsgedächtnisses beim Lernen quantitativer Schätzungen. In K. Bittrich, S. Blankenberger, & J. Lukas (Eds.). *Beiträge zur 53. Tagung experimentell arbeitender Psychologen* (p. 72). Lengerich: Pabst.

Hoffmann, J., & Marewski, J.N. (2010, March). Buy what you know: Einfache Heuristiken bei Konsumentenentscheidungen. In C. Frings, A. Mecklinger, D. Wentura, & H. Zimmer (Eds.). Beiträge zur 52. Tagung experimentell arbeitender Psychologen. (p. 150). Lengerich: Pabst.

Hoffmann, J., & Marewski, J.N. (2009, August). How can one get consumers to buy low-quality products? Poster presented at the 22nd Biennial Conference on Subjective Probability, Utility and Decision Making, Rovereto, Italy.

INVITED PRESENTATIONS

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). Pillars of judgment: How memory abilities affect rule-based and exemplar-based judgments. Thinking & Reasoning Group (Prof. S. Handley), University of Plymouth, United Kingdom.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2012). From simple rules to exemplar retrieval: Difficult tasks encourage shifting to memory-based strategies in judgment and categorization. Cognition & Individual Differences (Prof. E. Erdfelder, Prof. R. Pohl) and General Psychology (Prof. A. Bröder) labs, University of Mannheim, Germany.

ORGANIZED CONFERENCES & SYMPOSIA

Jenny, M., & Hoffmann, J. A, (March, 2013). Cognitive modeling in judgment and decision making. 55. Tagung experimentell arbeitender Psychologen, Vienna, Austria.
Participants: Marc Jekel, Thorsten Pachur, Nathaniel Philipps, Arndt Bröder, Mirjam Jenny, Janina Hoffmann

Hoffmann, J. A., & von Helversen, B. (September, 2012). Gedächtnisbasierte Entscheidungen. 48. Kongress der Deutschen Gesellschaft für Psychologie, Bielefeld, Germany.
Participants: Edgar Erdfelder, Anika K. Josef, Patrick, Khader, Christine Platzer, Janina Hoffmann, Bettina von Helversen

5th Workshop of Young Researchers in Judgment and Decision Making (2012), Basel, Switzerland

4th Workshop of Young Researchers in Judgment and Decision Making (2011), Bonn, Germany

SCHOLARSHIPS, GRANTS, & PRIZES

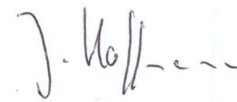|	
07/2012	5th Workshop of Young Researchers in Judgment and Decision Making sponsored by the University of Basel (2'690 CHF)
06/2012	SNF sponsored participation in the 2012 SNF Summer School in Computational Modeling of Cognition, Bergün, Switzerland (1'200 CHF)
11/2011	Student Poster Award of the Society of Judgment and Decision Making (250 US-$)
05/2010	Scholarship of the Graduate School of Economic and Social Sciences (GESS, University of Mannheim; declined).
2008/02 - 2010/02	Scholarship of the University of Mannheim (2'000 €)

MEDIA COVERAGE

Deliberation's blind sight: How cognitive load can improve judgments
Radio: Energie Basel
Newspapers: Blick am Abend, Aargauer Zeitung, Migros
Online: Spiegel Online

Basel, May, 30th, 2014

_____

Janina Hoffmann