von Helversen, Bettina ; Rieskamp, Jörg

Models of Quantitative Estimations: Rule-based and Exemplar-Based Processes Compared

Journal Article as: peer-reviewed accepted version (Postprint)

DOI of this document*(secondary publication): https://doi.org/10.26092/elib/3455

Publication date of this document: 08/11/2024

* for better findability or for reliable citation

Recommended Citation (primary publication/Version of Record) incl. DOI:

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(4), 867–889. https://doi.org/10.1037/a0015501

Models of Quantitative Estimations:

Rule-Based and Exemplar-Based Processes Compared

Bettina von Helversen & Jörg Rieskamp

Max Planck Institute for Human Development

Corresponding author's address:

Bettina von Helversen

Max Planck Institute for Human Development

Königin-Luise-Str. 5

D-14195 Berlin, Germany

Phone: +49 (0)30 8240 699

Fax: +49 (0)30 824 06 394

Email: vhelvers@mpib-berlin.mpg.de

Models of Quantitative Estimations:

Rule-Based and Exemplar-Based Processes Compared

The cognitive processes underlying quantitative estimations vary. Past research has identified task-contingent changes between rule-based and exemplar-based processes (Juslin, Karlsson, & Olsson, 2008). Von Helversen and Rieskamp (2008), however, proposed a simple rule-based model, the mapping model, that outperformed the exemplar model in a task thought to promote exemplar-based processing. This raised questions about the assumptions of rule-based versus exemplar-based models that underlie the notion of task-contingency of cognitive processes. Rule-based models, such as the mapping model, assume the abstraction of explicit task knowledge. In contrast, exemplar models should profit if storage and activation of the exemplars is facilitated. Two studies tested the importance of the two models' assumptions. When knowledge about cues existed, the rule-based mapping model predicted quantitative estimations best. In contrast, when knowledge about the cues was difficult to gain, participants' estimations were best described by an exemplar model. The results emphasize the task contingency of cognitive processes.

Keywords: decision making; simple heuristics; multiple cue judgments; quantitative estimation

Models of Quantitative Estimations:

Rule-Based and Exemplar-Based Processes Compared

How do people estimate a continuous quantity, such as the selling price of their house or the quality of a job candidate? In many cases people base their estimations on attributes or features of the object under evaluation that are probabilistically related to the quantity being estimated. For example, when estimating the selling price of a house people could rely on features such as the size of the house, the attractiveness of the neighborhood, or the presence of a deck. Cognitive models of estimation try to explain which features people use and how they integrate them to estimate a continuous criterion, that is, the quantity of interest.

Although quantitative estimation tasks are structurally quite similar to categorization tasks, two different research traditions have dominated in the two fields. In quantitative estimations, previous research has largely relied on linear additive models for describing people's estimations, such as multiple linear regression, while exemplar models prevailed in categorization. However, recently these two fields have been linked, and exemplar models originally proposed for categorization tasks have been introduced to the area of estimation (Juslin, Olsson, & Olsson, 2003). In this vein, Juslin, Karlsson, and Olsson (2008; Karlsson, Juslin, & Olsson, 2007) have argued that people frequently do not apply rules when making estimations but instead rely on an exemplar-based process. According to exemplar models people estimate the criterion of an object by retrieving the criterion values of similar exemplars from memory. Alternatively, von Helversen and Rieskamp (2008) have proposed the mapping model. The mapping model assumes that people estimate the criterion value of an object by first categorizing the object based on its features and then using a typical criterion value of past objects falling into the same category as an estimate. Although the exemplar model and the mapping model argue for conceptually different estimation processes, both models have been proposed for estimation tasks in which the standard regression approach did not provide a good account of people's estimations. Assuming that cognitive processes such as estimation are context dependent, the goal of the present article is to specify the judgment characteristics that lead to cognitive processes of quantitative estimations that can be best described by the two models. Furthermore, we aim to strengthen the connection between the two research traditions on categorization and estimation.

Models of Estimation

Consistent with the widespread assumption that human cognition comprises multiple processing modes (Ashby, Alfonso-Reese, Turken, &Waldron, 1998; Erickson & Kruschke, 1998; Hahn & Chater, 1998; Nosofsky, Palmeri, & McKinley, 1994), we assume that the cognitive processes underlying quantitative estimations can be described by distinct cognitive models. These models of estimation can be broadly classified by the underlying processes they assume, into rule-based models and more implicit, similarity-based models (Hahn & Chater, 1998; Juslin et al., 2003; Olsson, Enkvist, & Juslin, 2006; Patalano, Smith, Jonides, & Koeppe, 2001). Rule-based models rely heavily on the abstraction of information and often assume a controlled, serial, and verbally accessible cognitive process. In contrast, similarity- or exemplar-based processes rely on minimal abstraction, storing specific instances (Juslin & Persson, 2002). They are often thought to be parallel in nature and not verbally accessible (Juslin & Olsson, 2004).

The dominant approach to quantitative estimation, assuming a linear additive estimation process, falls into the category of rule-based models (Anderson, 1981; Brehmer, 1994; Brunswik, 1952; Hammond, 1955; Hammond & Stewart, 2001). Accordingly, estimation processes are conceptualized as a process of weighting and adding information that can be captured by a multiple linear regression. According to regression models, each object can be described on several dimensions, so-called cues, which are probabilistically related to the quantity of interest, the criterion that is being estimated. The specific features or characteristics by which an object is described on a cue dimension are the cue values (e.g., the cue "size" could have a cue value "large"). For dichotomous cue values we define the cue characteristic indicating a higher criterion value as a positive cue value and the cue characteristic indicating a lower criterion value as a negative cue value. Regression models assume that for each cue, the relation between cues and criterion is abstracted and explicitly represented as a cue weight; the judgment is then made by summing the weighted cue values (Cooksey, 1996; Doherty & Brehmer, 1997). Linear regression has been successfully applied to analyze judgments in many areas, such as clinical diagnostics (e.g., Harries & Harries, 2001), legal and medical decision making (Ebbesen & Konecni, 1975; Wigton, 1996), and personality evaluations (e.g., Zedeck & Kafry, 1977; for a review, see Brehmer & Brehmer, 1988).

However, even though linear additive models can capture human estimations quite well, in a linear environment, where the criterion is a linear function of the cues, they provide a less good description if the criterion is a nonlinear function of the cues (von Helversen & Rieskamp, 2008; Karlsson et al., 2008; Olsson et al., 2006). Further, regression models have been criticized for being "as if" models that often can describe the outcome of a decision but not capture the cognitive process underlying it (Gigerenzer & Kurz, 2001). In response to this criticism, von Helversen and Rieskamp suggested an alternative rule-based model, which they called the mapping model. This model can capture estimations in nonlinear environments.

The Mapping Model

The mapping model assumes a rule-based estimation process. Accordingly, people estimate the criterion value of an object by first categorizing the object and then assigning a typical criterion value of past objects falling into the same category as the estimate. The categorization process follows a simple rule, counting the number of relevant cues with a positive value. Each number of positive cue values constitutes a category. For example, when estimating the price of a house, the mapping model assumes that people consider the features of the house that favor a high price (e.g., great location, nice garden, a swimming pool). Then the number of positive features is used to categorize the house into a certain price class and the typical price for houses within this price class is used as an estimate.

The estimation process assumed by the mapping model is inspired by the framework for quantitative estimation developed by Brown and Siegler (1993). Brown and Siegler proposed that two types of information are necessary for an estimation: knowledge about the *mappings*, that is, the ordinal relation of the objects according to the criterion of interest; and knowledge about the *metrics*, that is, the numeric properties of the objects, such as the distribution, the range, or the mean of possible estimates. Both properties are necessary for an accurate estimation but are based on different sources of knowledge. The mapping model provides a computational account of Brown and Siegler's framework for estimation: In a first step, knowledge about the mappings is inferred from the cue values by counting the number of positive cue values and grouping objects together according to their cue sums. In a second step, knowledge about the metric properties is derived by abstracting a typical estimate for each category, represented by the median criterion values of the objects falling into the same category.

By categorizing objects according to the sum of positive cues, the mapping model assumes that all considered cues are weighted equally. Although counterintuitive at first glance, this assumption is based on robust findings showing that using varying weights for cues does not necessarily increase a model's predictive power (e.g., Dawes, 1979; Einhorn &

Hogarth, 1975; Hogarth & Karelaia, 2005) and that tallying models often provide a good description of human judgment processes (e.g., Bröder & Gaissmaier, 2007). However, it is important to note that the mapping model, in contrast to a unit weight regression model, allows the estimation of nonlinear cue–criterion relations, as the typical criterion values for each category are based on a central tendency of past objects.[1] Further, empirical results on the mapping model have shown that it predicts estimations in a variety of task environments quite well (von Helversen & Rieskamp, 2008). However, Juslin et al. (2008) recently proposed exemplar models as a competing account for nonlinear judgment tasks.

The Exemplar Model

In contrast to the mapping model, the exemplar model assumes a similarity-based process, according to which people estimate the criterion of an object by retrieving the criterion values of similar exemplars in memory. For example, when estimating the price of a house, the exemplar model assumes that people recall the selling prices of similar houses that were sold in the vicinity and use them to estimate the selling price for the house under evaluation. Exemplar models have been successfully employed to explain human behavior in categorization (Juslin et al., 2003; Kruschke, 1992; Nosofsky & Johansen, 2000). For categorization problems people classify objects into one of two or more categories based on the objects' features. Although categorization and quantitative estimation problems have a similar structure—the main difference being that estimations ask for a continuous judgment, while categorizations ask for a binary judgment—they have been studied by two different research traditions. For categorization problems similarity-based models such as exemplar or prototype models have predominantly been proposed, while for estimation problems linear additive models have been suggested. Only recently, connections between the two research fields have been established (e.g., Juslin et al., 2003) and exemplar models have been extended to the area of quantitative estimation (Juslin et al., 2003, 2008; Karlsson et al., 2007; Olsson et al., 2006). We think it is a fruitful approach to examine how cognitive models that have been proposed for one judgment domain can be generalized to another. When elaborating the connections between the research domains, the empirical findings from one domain, such as categorization, will be informative for our understanding of the cognitive process underlying judgments in another domain, such as quantitative estimation.

In general, exemplar models assume that estimations rely on the similarity of an object to previously encountered objects that are stored in memory. When applying exemplar models to estimation, it is assumed that previously encountered exemplars are activated and compared to the probe, that is, the object under evaluation. The more the probe resembles an activated exemplar, the closer the estimate for the probe will be to the exemplar's criterion value. More specifically, the estimate consists of the average criterion values of the activated exemplars, weighted by their similarity to the probe:

$$\hat{y}_p = \frac{\sum_{i=1}^{I} S(p,i) \cdot x_i}{\sum_{i=1}^{I} S(p,i)} \qquad (1)$$

where $\hat{y}_p$ is the estimated criterion value for the probe p; S is the similarity of the probe to the stored exemplars; x_i is the criterion value of the exemplar i; and I is the number of stored exemplars in memory. The similarity S between a stored exemplar and the probe depends on how many features the exemplar and the probe share. It is calculated using the multiplicative similarity rule of the context model (cf., Medin & Schaffer, 1978), defined as

$$S(p,i) = \prod_{j=1}^{J} d_j \, . \qquad (2)$$

For each cue *j* it is determined whether the cue values of the probe *p* and the stored exemplar *i* match. If they match, *d* equals one, and if they do not match, *d* equals the attention parameter s_j, which captures the impact of a cue on the overall similarity and varies between zero and one. The closer s_j is to zero, the more important the cue. If $s_j = 1$, this implies that the cue *j* is irrelevant for the evaluation of the overall similarity. The original exemplar model assumes a separate s_j parameter for each cue *j* (Juslin et al., 2003; Medin & Schaffer, 1978). However, as the original exemplar model seems to be prone to overfitting, we additionally considered a simplified version with one single attention parameter *s* for all cues (von Helversen & Rieskamp, 2008). In this case, *s* is an attention parameter indicating how closely a retrieved exemplar needs to resemble the probe to be considered for the estimation. The closer *s* is to zero, the more similar an exemplar has to be to the probe so that it has an impact on the estimation.

The above describe exemplar model illustrates how a model from the categorization domain is applied to the estimation domain. Naturally, exemplar models are not the only models that have been proposed for categorization processes. Another prominent approach to categorization is represented by prototype models. Prototype models assume that each category can be represented by one single prototype and that objects are categorized by their similarity to the prototype. It is an interesting question how and to what extent prototype models could be applied to the estimation domain. However, because it is not obvious how several components of prototype models could be specified for estimation problems, we consider this generalization as a research project on its own to be tackled in future research.

Model Environment Contingency

Both the exemplar model and the mapping model provide new and successful modeling approaches to quantitative estimation. However, both models were proposed to explain estimation processes in nonlinear estimation environments. Furthermore, two previous experimental studies led to rather conflicting result regarding which model provided a better account of observed estimations. In the third experimental study reported by von Helversen and Rieskamp (2008), the mapping model clearly outperformed the exemplar model in predicting participants' estimations. In contrast, a reanalysis of the first experiment of Juslin et al. (2008) as reported in von Helversen and Rieskamp revealed an advantage of the exemplar model over the mapping model in predicting estimations. Although the cover stories of the estimation problem differed, the structures of the two studies were very similar: In both studies, participants estimated a continuous criterion based on multiple dichotomous cues. The criterion was a multiplicative function of the cues, and the participants received outcome feedback to learn the task. However, the tasks differed in two aspects. First, in the study by Juslin et al. the training set was smaller and less complex and participants received twice as much training. Second, in the study by von Helversen and Rieskamp participants were informed about the direction of the cues, that is, which cue values indicated a high or a low criterion value. In contrast, in the study by Juslin et al., participants needed to learn the directions of the cues. Following the assumption that human cognition can be understood as an adaptation to different environments (Ashby & Maddox, 2005; Gigerenzer & Todd, 1999; Payne, Bettman, & Johnson, 1993; Rieskamp, 2006; Rieskamp & Otto, 2006), we propose in this article that the conflicting results found by von Helversen and Rieskamp (2008) can be explained by linking the characteristics of the task to the assumptions the models make about the estimation process. More specifically, we will first examine *exemplar memory*, that is, the ability to accurately store and activate previously encountered exemplars. Exemplar memory could be affected by the number of exemplars and the frequency with which the exemplars are encountered and consequently could influence whether an exemplar-based estimation processes becomes more likely. Second, we will investigate *prior knowledge* about the estimation task. Rule-based models require knowledge abstraction, that is, the knowledge that needs to be abstracted to execute a specific estimation process. *Prior knowledge* about an estimation task could make this abstraction process easier and thereby foster a rule-based estimation process that requires specific abstracted knowledge.

Exemplar memory. Exemplar models assume the activation of previously encountered exemplars from memory. If an individual is able to activate encountered exemplars easily and accurately, then relying on an exemplar-based estimation process should be accurate and might even become cognitively less demanding than a rule-based process. Due to these reinforcements an exemplar-based estimation process could become more likely. Research on list learning has shown that recognition and recall improve (1) the fewer the items to be learned and (2) the more frequently they are repeated during training (e.g., Gillund & Shiffrin, 1984). Similarly, a lower number of training exemplars and a higher frequency of repetitions of exemplars could foster an exemplar-based estimation process and consequently increase the accuracy of exemplar models in predicting people's estimations.

In this vein, in the area of categorization it has been suggested that exemplar-based processes are more prevalent for small categories with few dimensions and easily distinguishable exemplars (Ashby & Ell, 2001; Minda & Smith, 2001; Rouder & Ratcliff, 2006). Similarly, Homa, Proulx, and Blair (2008) proposed that more exemplars in a category lead to the abstraction of prototypes while categories with fewer exemplars trigger exemplar-based processes. Similar to prototype models, the mapping model assumes a high degree of abstraction: Instead of memorizing exemplars with all their feature values, memory load is reduced to category membership, making it less sensitive to the number of training exemplars. Thus, while a low number of training exemplars might foster an exemplar-based process, a large number of training exemplars might—analogous to the prototype abstraction process assumed for large categories (Homa, Dunbar, & Nohre, 1991; Homa et al., 2008)—favor a rule-based estimation process as described by the mapping model. Accordingly, the mapping model should predict estimations better than the exemplar model in a situation with a large number of training objects in comparison to a situation with a smaller number of training objects. This *exemplar memory prediction* will be tested in Study 1.

However, the proposed relation between exemplar memory, that is, the effect of category size and frequency of exemplars on the reliance on exemplar-based processes, has been contended by proponents of exemplar theory and is widely discussed in the literature (e.g., Knowlton & Squire, 1993; Nosofsky, 1988a, 1988b; Nosofksy & Zaki, 1998; Shin & Nosofsky, 1992). In particular, multiple trace approaches to exemplar memory—although they assume that the probability and accuracy with which an exemplar is recognized increases with the frequency of presentation (Hintzman, 1988)—would not necessarily expect that the number of training exemplars affects the judgment process. Smith and Minda's (1998; Minda & Smith, 2001) results of examining the effect of the number of training exemplars on exemplar-based processes have also been challenged by various researchers (Nosofsky & Zaki, 2002; Olsson, Wennerholm, & Lyxzen, 2004; Zaki, Nosofsky, Stanton, & Cohen, 2003). In sum, there are mixed findings and views in the area of categorization processes about whether the ability to retrieve exemplars has an impact on the categorization process. More crucially, its role in the area of quantitative estimations has not been considered systematically at all, which is a novel contribution of our research. The ongoing debate suggests that it is worthwhile to consider exemplar memory as a possible factor influencing estimation processes.

Prior knowledge. Differences in the availability of task knowledge could also explain the diverging results reported by von Helversen and Rieskamp (2008) and Juslin et al. (2008). While the exemplar model might rely on accurate storage and activation of encountered exemplars, the mapping model requires the abstraction of explicit task knowledge. In general, rule-based processes, such as the one described by the mapping model, require specific knowledge about the judgment task, which is then represented in an abstract, explicit form. For instance, when applying a rule for categorization the rule needs to abstract thresholds that allow accurate categorizations (e.g., Ashby et al., 1998; DeCaro, Thomas, & Beilock, 2008) and are then explicitly accessible. Ashby and colleagues showed that rule-based categorization processes were primarily followed when the stimulus dimensions were separable. However, if the stimuli were integral, similarity-based processes were relied upon. This suggests that the ease with which an individual can abstract necessary knowledge when encountering the judgment task could affect the likelihood with which a rule-based process will occur. In particular, prior knowledge about a task should make a rule-based process that requires this knowledge more likely.

In the third experimental study of von Helversen and Rieskamp (2008) participants were informed about the directions of the cues, providing explicit, verbally accessible knowledge about the estimation task. However, in the study by Juslin and colleagues (2008), no prior information about the cues was given to the participants, making it more difficult to gain knowledge of the cue directions. For the mapping model, abstracting knowledge about the predictability and the directions of the cues is crucial to grouping objects into meaningful categories. Thus, if prior knowledge about the cue directions exists, this could foster rule-based processing consistent with the mapping model, as participants do not need to invest resources into learning the relationship between the cues and the criterion and can quickly achieve a good level of performance. In contrast, the exemplar model, relying on the similarity relations of the objects, does not depend on knowledge about the cue directions but can be applied successfully as long as objects are sufficiently differentiable (although it also needs to learn which cues are relevant). In particular, if cue directions are difficult to learn, it might be more demanding to gain knowledge about the cue directions, needed to sort the objects into the correct categories, than to rely on exemplar memory. Therefore, if no prior knowledge about the cues exists, this could trigger exemplar-based processing.

In sum, rule-based estimation processes as described by the mapping model rely on the abstraction of knowledge about the cues and should profit more than exemplar-based processes from prior knowledge about the task, such as information about the cues' directions. From this, the *prior knowledge prediction* follows: rule-based processes should be more likely observed when the necessary knowledge is easily attainable. In contrast, the exemplar-based processes seems to be particularly suited to capturing the estimation process when it is difficult to abstract knowledge about the task necessary for a rule-based process, but the memorized exemplars allow an accurate estimation performance (see also Olsson et al., 2006). We tested these predictions in two studies, manipulating the exemplar memory as well as the access to prior knowledge about the cues.

Methods of Model Selection and Qualitative Tests of Models

Model selection can be a challenging task. For one thing, model complexity needs to be taken into account: More complex models are better in fitting data but they run the risk of overfitting; that is, they not only capture systematic variance but also fit unsystematic variance in the data. Another challenge is that models often make very similar predictions, making it difficult to devise tests that reliably differentiate between the models.

We addressed the problem of model selection with a twofold approach. First, we compared the models on the basis of a generalization test (Busemeyer & Wang, 2000). Generalization tests follow the rationale that if a model captures the cognitive process underlying estimations then this model should also predict new independent estimations better than alternative models that do not capture the cognitive process. They not only allow a fair comparison of models of different complexity but go beyond a pure cross-validation test, providing a rigorous model-selection method. We implemented the generalization test by first estimating the models' parameters using the data of a training phase and then using the estimated parameter values to predict participants' estimations for new test objects.

Second, we devised a qualitative model comparison test. Qualitative tests are preferable to purely quantitative model tests (Pitt, Kim, Navarro, & Myung, 2006). They are less dependent on specific parameter values and they provide a critical test of the model assumptions, providing information about the correspondence of the pattern in the data with model predictions. Therefore, we aimed to find qualitative predictions that were specific for each model and could not be derived from the competing model.

For this purpose we focused on the assumptions the models make about which objects should be treated similarly and for which objects the estimations should differ. The mapping model groups objects according to their cue sums, ignoring which specific cues have a positive value. This implies that the model makes the same estimations for all objects with the same cue sum, whereas estimations for objects with different cue sums will differ. The exemplar model, on the other hand, relies on the similarity relations of the objects to the stored exemplars. Thus two objects that are maximally different should also differ in which exemplars they resemble and thus in the criterion values estimated. We used these assumptions of the models to design qualitative model comparison tests in addition to the quantitative model comparison tests.

Study 1

The goal of Study 1 was to investigate the influence of exemplar memory on model performance for quantitative estimations. This seems to be an important question, as the role of an accurate representation of exemplars in memory has been proposed as a potentially important factor influencing categorization processes but has not been considered at all for the area of quantitative estimations. Moreover, when comparing the different models we go beyond the pure quantitative tests provided by von Helversen and Rieskamp (2008) and test models against each other by focusing on qualitative criteria, which should allow a better discrimination between the models.

In the multiple-cue estimation task of the experimental study the participants evaluated job candidates based on six cues. In a training phase participants were presented with a number of candidates and the evaluations they had received. Based on this training sample they could learn how their company evaluated job candidates. In a subsequent test phase, we tested how well they generalized this knowledge to new job candidates and which model was best in predicting their evaluations. Study 1 focused on the exemplar memory prediction described above. Accordingly, we manipulated the number of training objects in two experimental conditions, examining whether less reliance on exemplar-based processes in favor of a rule-based process as described by the mapping model could be observed with a larger number of training objects.

Method

Participants. In Study 1, 40 participants took part, 20 in each condition. The majority of participants were students from one of the Berlin universities, with an average age of 24 years (*SD* = 4); 30% of the participants were male. Participants were randomly assigned to one of the experimental conditions, balanced for gender. The study lasted for about 1 h 45 min and participants were paid an average of €16 for their participation. One participant in the condition with a small number of training exemplars was excluded from the analysis as he did not improve in evaluating the training candidates during the training phase.

Procedure and material. The study was conducted as a computer-based experiment. The task of the participants was to evaluate the quality of job candidates for an Information Technology (IT) position on a scale of 1 to 100 points. The more points a job candidate received the more suited he or she would be for the position. Participants received information about the job candidates on six dichotomous cues. The six cues and their characteristics were knowledge of programming languages (C++ vs. Java), knowledge of foreign languages (French vs. Turkish), additional skills (SAP, a software system, vs. web design), previous work experience (software development vs. system administration), previous employment area (business vs. academia), and knowledge of operating systems (UNIX vs. Windows).

Participants were told which of the two possible characteristics of the cues matched the company's needs; characteristics that matched the company's preferences were marked in green, while characteristics that did not meet the company's requirements were marked in red. During training participants learned how many points job candidates with different combinations on the six cues should receive. The criterion, that is, how many points a job candidate received, was determined as a multiplicative function of the cue values (von Helversen & Rieskamp, 2008; Juslin et al., 2008):

$$C = 0.68 \cdot e^{(22c_1 + 20c_2 + 17c_3 + 15c_4 + 14c_5 + 12c_6)/20}, \tag{3}$$

where C is the points the job candidate received and c_1 to c_6 are the values on the six dimensions. A positive characteristic of a cue was coded with a cue value of 1 and a negative characteristic was coded with a cue value of zero. The assignment of the weights to the cues, which characteristic of a cue was coded as positive or negative, as well as the order of the cues on the screen were randomly determined for each participant.

The study consisted of two parts, a training phase and a test phase. During the training phase participants could learn the company's evaluation policy by judging job candidates. In each trial participants were asked to evaluate one job candidate. After each trial they received feedback about the number of points this candidate should receive, how close their estimate had been, and how many points they earned in this trial (see below). Then the next candidate appeared. All training candidates were repeated 10 times, structured in 10 blocks; the order of appearance in each block was randomly determined.

We manipulated the number of training candidates in this study: In one condition the training set consisted of a large number (24) of different training candidates; in the other condition the training set comprised a small number (8). After the training phase participants continued with a test phase in which they had to evaluate 30 more job candidates. The test phase was similar to the training phase, with the difference that participants did not receive immediate feedback about the accuracy of their evaluations and only learned how many points they had earned after they had finished the test phase. The 30 test candidates were evaluated twice. Eight of the candidates in the test phase had also appeared during training and 22 were new candidates participants had not encountered before.

Participants' payment was based on their performance. In each trial participants could earn up to 100 points depending on how accurately they estimated the quality of the job candidates. The more they deviated from the criterion the fewer points they earned. The exact number of points subtracted for a given deviation was calculated with a feedback algorithm, based on the squared deviation from the estimation to the criterion.[2] After the experiment, points were exchanged into euros at a rate of €0.1 for every 150 points.

Selection of training and test sets. To test which model could explain the participants' behavior best we relied on a generalization test (see Busemeyer & Wang, 2000). Additionally we conducted a qualitative test of the models' assumptions (see Pitt et al., 2006), focusing on two qualitative predictions that were derived from the models' assumptions about the estimation process.

First, according to the mapping model, the same value is estimated for any two objects with an equal number of positive cues, regardless of the similarity of the two objects. In contrast, if two objects are very dissimilar, that is, if they do not match on a single cue, the exemplar model's estimations should differ. For the experimental task with six cues, an estimation situation like this occurs for objects with a cue sum of three. To clarify, for any cue profile with three positive and three negative cues (e.g., 111000, with each number representing the cue value of one cue), the mapping model makes the same prediction for an object with the reversed cue profile (i.e., 000111). In contrast, the exemplar model will most likely make different estimation predictions, because these two objects are maximally dissimilar.

Second, we devised an additional experimental situation in which the exemplar model and the mapping model made opposite predictions. The mapping model predicts large differences between estimates for objects when they have different cue sums, for instance, objects with cue sums of 2 and 4. In contrast, for these objects, which necessarily share some cue values, the exemplar model can make very similar estimations.

To summarize, our qualitative test comprised two conditions in which the exemplar model and the mapping model made qualitatively different predictions. While the mapping model predicted a difference between the estimations for objects with a cue sum of 4 and a cue sum of 2 and no difference for objects with a cue sum of 3, the exemplar model made the opposite predictions. However, the strength of the qualitative predictions depends on the specific training and test objects. For instance, if all training objects had the same criterion value, it would be impossible to differentiate between the models. Accordingly, we aimed at selecting training set–test set combinations where the qualitative predictions of the two models would differ as widely as possible, but we made sure that the test set would well represent the total set. The test set for the condition with a large training set is reported in Table 1. For the training sets, the test set of the condition with eight objects, and details on the procedure used to select the training and test sets, please see Appendix A.

Finally, we explored the prediction of the models for the models' parameter space, to determine the range of parameter values for which the models make qualitatively different predictions. For the mapping model this is a simple task because it has no free parameters, so it makes one single prediction for a specific object. In contrast, the exemplar model's predictions depend on its values for the attention parameter.[3] We covered the parameter space of the attention parameter *s* by using the values .001, .1, .2, .5, .7, and .9. Figure 1 illustrates how the predictions of the exemplar model change with increasing parameter values. With small parameter values a clear difference in the qualitative predictions of the models can be observed. The small values for the attention parameter of the exemplar model are most plausible, because they are exactly the ones von Helversen and Rieskamp (2008) found to be the best estimates for the exemplar model (i.e., the average estimated parameter values varied between .001 and .17; see also Juslin et al., 2008, for similar low parameter estimates for the standard exemplar model). Thus, when assuming small attention parameter values that perform best in predicting participants' estimations, the two models make distinct ordinal predictions. Moreover, the results show that over the whole range of parameter values, the models' predictions do not overlap and that even for parameter values for which the exemplar model predicts the same ordinal data pattern as the mapping model, strong quantitative differences are to be expected.

Results

Overall, the mapping model predicted participants' estimations significantly better than the exemplar model in both conditions. The advantage of the mapping model was higher in the condition with a small number of training objects than in the condition with a large number of training objects, indicating that contrary to the exemplar memory prediction the number of training exemplars does not lead to a more rule-based estimation process. However, before we come to the model comparisons, we first report participants' accuracy.

Participants' accuracy. Participants learned to evaluate the training candidates fairly well in both conditions. We measured the participants' accuracy via the root mean square deviation (*RMSD*) between the criterion values and the participants' estimations. In the condition with a large number of training objects *RMSD* dropped from 15.56, $SD = 5.62$ in the first block to 3.86, $SD = 2.07$ in the 10th block. Similarly, the *RMSD* in the condition with a small number of training objects dropped from 22.97, $SD = 6.76$ in the first block to 3.04, $SD = 4.04$ in the 10th block. Because the participants' performance as well as the model fits were not normally distributed we used nonparametric tests (the Mann–Whitney *U*-test for independent samples and the Wilcoxon *Z*-test for paired samples) throughout the article. The participants' accuracy in the test phase did not differ between the two conditions, $RMSD_{\text{large}} = 5.84$, $SD = 1.87$ versus $RMSD_{\text{small}} = 7.42$, $SD = 3.39$; $U = 137$, $p = .14$. However, in both conditions, accuracy in the test phase was worse than in the training phase, $RMSD_{\text{training}} = 3.98$, $SD = 2.32$ versus $RMSD_{\text{test}} = 6.61$, $SD = 2.80$; $Z = -4.16$, $p < .01$. Participants were more accurate in the test phase in estimating the old objects known from the training phase than the new objects, $RMSD_{\text{old}} = 4.49$, $SD = 6.10$ versus $RMSD_{\text{new}} = 6.69$, $SD = 2.21$; $Z = -3.99$, $p < .01$.

Overall, participants were quite consistent in their estimations. Consistency was measured as the Pearson correlation between the first and the second presentation of the test objects. In both conditions consistency was similarly high, $r_{\text{large}}(20) = .95$, $SD = .06$ versus r_{small} (19) = .94, $SD = .06$; $U = 137$, $p = .14$. Overall, participants were more consistent in estimating old objects than new objects, $r_{\text{old}} = .98$, $SD = .05$ versus $r_{\text{new}} = .85$, $SD = .15$; $Z = -4.88$, $p < .001$.

Model parameters. To test which model predicted participants' estimations best, we first fitted both models on the last blocks of the training phase for each participant individually. In the condition with a large number of training objects we used the last three blocks and in the condition with a small number of training objects the last four blocks to fit the models on a sufficient number of training objects. Based on the parameters estimated we made predictions for the test phase. Goodness-of-fit was determined as the *RMSD* of the model prediction from the participants' estimations. For the exemplar model we fitted two versions: the standard exemplar model with a free attention parameter for each cue and the simplified exemplar model with a single free parameter. The exemplar model's parameters were estimated by using participants' estimations for the last blocks of the training phase with a knowledge base consisting of the objects from the training phase with their correct criterion values. The best values for the model's free attention parameters were found by a grid search followed by a nonlinear least square method (as implemented in MATLAB). For the mapping model no parameters needed to be estimated. We simply calculated the typical criterion value (i.e. the median) for all objects of the training set with the same cue sum (using the correct criterion values of the objects).

In addition to comparing the exemplar model and the mapping model, we also tested further linear additive models to rule out that they predicted participants' behavior better. More specifically, we included a multiple linear regression model (Brehmer, 1994; von Helversen & Rieskamp, 2008; Juslin et al., 2008) and a simpler unit weight model (Dawes, 1979). Lastly, one might object that a complex model such as the standard exemplar model might be at a disadvantage in a generalization test, because its parameters are estimated on a training sample that differed from the test sample. Although we do not agree with this objection, because a model should be able to make worthwhile predictions for new independent estimation situations, we additionally conducted a cross-validation test. The results of this test were consistent with the main generalization test and its results are reported in Appendix B.

Quantitative model comparison. In the training set both models described participants' estimations fairly well (for means see Table 2). For the training phase the standard and the simplified exemplar model performed better than the mapping model in both conditions (mapping vs. simplified exemplar model: $Z_{\text{small}} = -2.20$, $p = .03$; $Z_{\text{large}} = -3.21$, $p < .01$) and the linear additive models (regression vs. simplified exemplar model: $Z_{\text{small}} = -3.83$, $p < .01$; $Z_{\text{large}} = -3.92$, $p < .01$). However, the better fit of the exemplar model during training can be explained by its higher flexibility due to its free parameter(s) and should not be decisive for model selection. The crucial test is how well the models predict participants' estimations in the test phase for the new objects they did not encounter during training.

Here, the mapping model was clearly the best model, outperforming the other models in both conditions. In the condition with a large number of training objects it reached a *RMSD* of 5.87, $SD = 2.32$, compared to the simplified exemplar model with a *RMSD* of 15.45, $SD = 2.37$; $Z = -3.92$, $p < .01$ or the standard exemplar model with a *RMSD* of 16.88. Also in the condition with a small number of training objects the mapping model ($RMSD = 5.74$, $SD = 3.52$) was clearly superior to the exemplar model ($RMSD_{\text{simplified}} = 22.63$, $SD = 1.82$; $Z = -3.82$, $p < .01$; see also Table 2). The mapping model also outperformed both linear additive models (the regression model: $Z_{\text{large}} = -3.92$, $p < .01$ and $Z_{\text{small}} = -3.82$, $p < .01$; and the unit weight model: $Z_{\text{large}} = -3.92$, $p < .01$ and $Z_{\text{small}} = -3.82$, $p < .01$). In general, the rule-based models performed better than the exemplar models. In the condition with the large number of training objects the regression model was the second best model; in the condition with the small number of training objects the unit weight model was second best.

Qualitative model comparison. Though the quantitative model comparison already showed that the mapping model was better at predicting participants' estimations than the exemplar model, we additionally relied on a qualitative test. The qualitative test was designed to specifically test the models' assumptions about the cognitive process underlying estimations. As the simplified exemplar model performed better than the standard exemplar model we compared the mapping model to the simplified version. To test the models' predictions, we determined for each participant and model the mean difference between the estimations for the objects with a cue sum of 2 and 4 and for the pair of objects with a cue sum of 3. As expected from the parameter space analysis illustrated in Figure 1, for both experimental conditions the models made clearly distinct qualitative predictions, as illustrated in Figure 2.

In the condition with a small number of training objects, the exemplar model predicted a small difference of 1.2 points while the mapping model predicted a difference of 22 points for test objects with cue sums of 2 and 4. In contrast, for the pairs of objects with a cue sum of 3, the mapping model predicted no difference, while the exemplar model predicted that estimations would differ by 18.4 points. Although not quite as pronounced, the same interaction was predicted in the condition with a large number of training objects. The predictions of the mapping model were clearly supported by the data. In both conditions participants' estimations differed strongly for the objects with a cue sum of 4 and a cue sum of 2 (M_{small} = 18.1 points, *SD* = 4.5 and M_{large} = 17.2 points, *SD* = 5.3), close to the difference predicted by the mapping model. Likewise, the participants' estimations for the objects with the same cue sum but maximally different cue profiles corresponded to the assumptions of the mapping model, M_{large} = 1.3, *SD* = 2.1 and M_{small} = −1.8, *SD* = 3.2.

Discussion of Study 1

Study 1 supported the mapping model in an estimation task with multiple predictive cues and a nonlinear cue–criterion relationship. The model predicted well how participants estimated values for objects they had not seen during training, obviously capturing the process underlying the estimations. Furthermore, it outperformed the exemplar model as well as a linear regression. In comparison, the exemplar model performed quite poorly; although it was able to accurately describe the estimations during training, it could not predict the estimations for the test phase and performed worse than the linear additive models. However, although the exemplar model failed to predict estimations for new items, it captured the estimations' for old items quite well, even outperforming the mapping model in the condition with a large set of training objects. This could indicate that participants employed different strategies for old and new objects, relying on memorization for old objects, but following a rule-based estimation process as described by the mapping model for the new objects (for a similar discussion see Rehder & Hoffman, 2005). However, the advantage for the exemplar model for old items might also be a modeling artifact. Due to its high flexibility the exemplar model was able to fit the estimations very well for the training phase. This good fit is also reached by taking inter-individual differences into account, which the mapping model ignores. Given the high consistency of participants' estimations, the estimations in the test phase did not differ substantially for the old items and consequently the exemplar model also reached a good fit for these items.

In contrast to the exemplar memory prediction that the exemplar model's predictions should improve with fewer training objects, the model performed worse, $U = 5$, $p <.01$. However, the implication of these results for this prediction is limited, because the mapping model outperformed the exemplar model in both conditions. This indicates that participants did not rely on exemplar-based processes in either condition, but that the difference in the performance of the exemplar model could also be due to modeling issues, for example, the increased chance to average out unsystematic variance with more training instances.

Thus, overall, the results indicate that the number of training objects is not a crucial factor affecting people's estimation processes. However, the mapping model could have outperformed the exemplar model because we provided knowledge about the cue directions. Research in categorization has reported strong effects of prior knowledge on cognitive processing (e.g. Wisniewski & Medin, 1994). Furthermore, according to the *prior knowledge prediction*, rule-based processes should be more likely to be observed when the necessary knowledge is easily attainable. Thus, prior knowledge about the cue directions could have triggered rule-based processing in accordance with the mapping model, overriding any effects of the exemplar memory prediction.

On the other hand, even though prior knowledge has been reported to affect cognitive processing in categorization (Kaplan & Murphy, 2000), this is much less clear for estimation. Furthermore, the exemplar model could solve the task perfectly without knowledge about the cue directions and past research has shown that people often ignore information even if it is relevant to the task (e.g. base rate neglect, Tversky & Kahneman, 1994; see also Payne et al., 1993). Thus, the lack of an effect of the number of training objects could also have been caused by too little training. Multiple systems theories of categorization often assume that people start with rule-based processes and use exemplar knowledge as a back-up strategy (e.g., Karlsson et al., 2008; Nosofsky et al., 1994; Rehder & Hoffman, 2005; Smith & Minda, 1998). Similarly, Johansen and Palmeri (2002) suggested a representational shift from rule-based to exemplar-based processes with extensive training, and Rouder and Ratcliff (2006) argued that the memorization of complex exemplars might take time and that people might rely on rule-based strategies until they have gained enough experience with the exemplars to store those that are immune to forgetting and then rely on the more accurate exemplar-based strategy. In our study every training object was repeated 10 times, leading to a quite accurate performance of the participants in the estimation task. Nevertheless, studies investigating exemplar-based approaches often provide more training. For instance, Zaki et al. (2003) presented training objects 40 times each and Juslin et al. (2008) presented each object 20 times. Thus the exemplar memory prediction might only hold when a higher amount of training is provided. We conducted Study 2 to examine the relevance of the length of training and prior knowledge for cognitive processing for quantitative estimation.

Study 2

In Study 1 the participants did not follow an exemplar-based estimation process. In Study 2 we addressed two possible reasons for the poor performance of the exemplar model in Study 1. Exemplar-based processes might be more likely to occur when extensive training is provided (Johansen & Palmeri, 2002). Thus, we increased the training to 20 blocks, similar to the study by Juslin et al. (2008). Second, the availability of explicit knowledge about the cues could have primed rule-based processing in Study 1. Because the exemplar model's performance is largely independent of explicit task knowledge, providing no information about the cues should create conditions favorable for the exemplar model. However, a shift to exemplar-based processing might depend not only on the availability of knowledge, but also on the ease with which knowledge can be gained. If picking up the cue directions during training is easy, the mapping model could still prevail. In Study 1 (see Table 3) all cues correlated substantially with the criterion, which should make it fairly easy to pick up the cues' directions, as learning of contingencies depends to a high degree on the strength of the relation (e.g., Brehmer, 1973; see also Hoffman & Murphy, 2006; Klayman, 1988a). Therefore, to vary the ease with which the cue directions could be learned, we also manipulated how demanding it was to detect the correct directions of the cues. For this purpose we created a training set in which only half of the cues were predictive whereas the other half were useless for estimating the criterion values. This should increase the difficulty of inferring the cues' directions for predicting the criterion (Brehmer, 1973).

Method

Participants. In Study 2, 80 students from one of the Berlin universities participated (average age = 25 years, *SD* = 3); 33% of the participants were male. Participants were randomly assigned to one of the four experimental conditions, balanced for gender. The study lasted for about 1 h 30 min and participants were paid on average €14 for their participation.

Design, procedure, and material. In Study 2 we increased the training phase, providing twice as many learning trials in comparison to Study 1. In addition, we manipulated the prior knowledge about the directions of the cues and the ease with which the cues' directions could be learned with two between-subjects factors in a 2 × 2 experimental design. Similar material to Study 1 was used. Again, participants were asked to evaluate the quality of job candidates based on the six binary cues described in Study 1. However, in Study 2 only half of the participants were told which cue values were regarded as positive and which as negative. The other half needed to discover the cues' directions during the training phase. Additionally, we manipulated how easily the cues' directions could be learned. One half of the participants were provided with the identical set of training objects used in the training phase of the condition with eight training objects in Study 1. For this set of training objects all cues correlated substantially with the criterion (in all cases $r > .35$). For the other half of participants we used a different set of training objects, so that three cues correlated highly with the criterion ($r > .5$) and three correlated poorly ($r < .2$). The exact cue–criterion correlations are reported in Table 3. The selection of objects for the training and test phases for the second condition was achieved in the same way as in Study 1 with the additional constraint on the cue–criterion correlations and the exclusion of extreme profiles (with all positive or all negative cue values, which had to be excluded to achieve the desired cue–criterion correlations).

Similar to Study 1, Study 2 consisted of a training phase and a test phase. The training sets in both conditions consisted of eight training exemplars. In comparison to Study 1, we increased the duration of the training to 20 trials per candidate, structured in 20 blocks. In each block the eight training candidates were presented in a random order. Participants were paid contingent on their performance, based on the same feedback algorithm used in Study 1.[4] The subsequent test phase consisted of 30 objects with 22 new and 8 old objects that participants evaluated twice. The test objects were selected in the same way as in Study 1 to allow a qualitative test of the models. The training and test sets are reported in Appendix A (Tables A2 and A3). After the test phase, participants who had not been informed about the cue directions were asked to indicate which cue values went with higher criterion values.

Results

As in Study 1, the mapping model outperformed the exemplar model when the directions of the cues were known to the participants. However, when the cue directions had to be learned during training, which model predicted the participants' estimations best depended on the number of predictive cues, that is, cues that correlated substantially with the criterion. In the condition in which all cues were predictive, the mapping model was still best in predicting the estimations. Only in the condition in which the directions of the cues were unknown to the participants and only three cues were predictive did the exemplar model outperform the mapping model.

Participant performance. The participants learned to evaluate the job candidates correctly in all conditions, dropping from an average *RMSD* of 27.31, *SD* =12.61 in the first block to 3.77, *SD* = 5.90 in the 20th block. However, training accuracy depended on the knowledge of the cue directions. Participants were more accurate in their estimations when they knew the cue directions (*RMSD* = 2.07, *SD* = 2.03) than when they did not (*RMSD* = 7.43, *SD* = 6.79; $U = 364$, $p < .01$). If the cue directions were known, participants did better if all cues were predictive (*RMSD* = 1.40, *SD* = 2.03) than if only half were predictive (*RMSD* = 2.75, *SD* = 1.82; $U = 94$, $p < .01$). However, when the cue directions were not known, participants performed equally well ($RMSD_{\text{three predictive cues}} = 6.11$, $SD = 4.09$ vs. $RMSD_{\text{six predictive cues}} = 8.74$, $SD = 8.62$; $U = 193$, $p = .86$). Overall, participants' estimation accuracy was better for the training phase than for the test phase ($RMSD_{\text{training}} = 4.75$, $SD = 5.66$ vs. $RMSD_{\text{test}} = 11.82$, $SD = 5.79$; $Z = -7.62$, $p < .01$).

To measure the consistency of participants' estimations we calculated the Pearson correlation between the two judgments of the same objects during the test phase. A similar pattern to that found for participants' accuracy emerged: Participants were more consistent when they knew the cue directions ($r = .92$, $SD = .11$) than when they learned them during training ($r = .81$, $SD = .17$; $U = 448$, $p < .01$). When the participants knew the cue directions, the number of predictive cues did not matter ($r_{\text{three predictive cues}} = .92$, $SD = .11$ vs. $r_{\text{six predictive cues}} = .92$, $SD = .10$, $U = 193$, $p = .86$). However, when the cue directions were learned during training, participants were more consistent when all cues were predictive ($r = .86$, $SD = .15$) than when only three cues were predictive ($r = .76$, $SD = .17$, $U = 122$, $p = .04$). Overall, participants were more consistent in estimating the old objects than estimating the new objects ($r_{\text{old}} = .93$, $SD = .14$ vs. $r_{\text{new}} = .79$, $SD = .22$; $Z = -5.50$, $p < .01$).

Knowledge of cue directions. To examine whether our manipulation of the ease with which the cue directions could be learned had an effect on the estimations, we compared how many mistakes participants made in reporting the correct directions of the cues. As expected, participants performed better when all six cues were predictive (i.e., correlated substantially with the criterion) than when only three cues were predictive. When all cues were predictive, 7 (35%) participants indicated for at least one cue an incorrect direction; whereas when only three cues were predictive, 14 (70%) participants made at least one mistake. In particular, the participants had difficulty in correctly reporting the direction of the low-quality cues (i.e., those that correlated only slightly with the criterion), with a total of 16 mistakes in comparison to only 8 mistakes with the high-quality cues.

Quantitative model comparison. As in Study 1 we used the last four blocks of the training phase to estimate individually the exemplar models' attention parameter. We used the objects' correct criterion values in the training phase to determine the median estimates for the mapping model's estimation categories. The categories were formed on the basis of all six cues.[5] In this way we determined the models' predictions for the new objects in the test phase. Model performance was measured as the *RMSD* between model predictions and participants' estimations. For the generalization test we focused on the simplified exemplar model, as it performed better than the standard exemplar model in Study 1. However, we provide the results of a full model comparison including the standard exemplar model, a multiple linear regression, and a unit weight model in Appendix B. Additionally we report the results of a cross-validation test when different from the results of the generalization test (for details on the cross-validation test see Appendix B).

As in Study 1, the generalization test focuses on how well the two models predict participants' estimations for the new independent objects of the test phase. Table 4 reports the mean *RMSD*s and *SD*s. Figure 3 shows that in the condition replicating the Study 1 condition with a small number of training objects (where the participants knew the cue directions and where all cues were predictive), with the only difference being having a larger number of training trials, the mapping model again clearly outperformed the exemplar model, $Z = -3.92$, $p < .001$. Thus, contrary to the exemplar memory prediction, simply having more training did not cause the participants to switch to an exemplar-based estimation process. Similarly, when the cue directions were known but only half of the cues were predictive, the mapping model predicted participants' estimations better than the exemplar model, $Z = -3.92$, $p < .01$. Furthermore, the mapping model was still the superior model when the participants had to learn the directions of the cues, and all cues were predictive, $Z = -2.80$, $p < .01$. However, in line with the knowledge abstraction prediction, when the participants needed to abstract the directions of the cues during training but only three cues were predictive, making abstraction of the necessary knowledge for the mapping model difficult, the exemplar model outperformed the mapping model, $Z = -3.62$, $p < .01$.

Overall, the cross-validation test led to similar results. However, the standard exemplar model performed better in the cross-validation test ($RMSD = 11.92$, $SD = 4.49$) than in the generalization test (*RMSD* = 21.53, *SD* = 10.34). In particular, in the condition with six predictive cues and no knowledge about the cue directions, the standard exemplar model performed as well as the mapping model ($Z = -.24$, $p = .84$) in cross-validation, with 9 participants better predicted by the exemplar model and 11 participants by the mapping model. Whether the mapping model outperformed the standard exemplar model depended on the participants' success in learning the cue directions: For 77% of the participants who did learn all cue directions correctly, the mapping model predicted the estimations better than the standard exemplar model, whereas the mapping model only did better than the exemplar model for 14% of the participants who did not learn all cue directions; χ^2 (2, N = 20) = 7.21, p = .02.

Qualitative model comparison. Similar to Study 1, we also tested which of the qualitatively different predictions of the two models were in line with the observed estimations. Again, we compared the predictions of the exemplar model and the mapping model by taking the difference in estimations for the pairs of objects with a cue sum of 3 and the objects with cue sums of 2 and 4. For the pairs of objects with a cue sum of 3 the mapping model predicted no difference between the estimates whereas the exemplar model predicted a large difference. In contrast, for the objects with cue sums of 2 and 4 the mapping model predicted a large difference and the exemplar model predicted a small difference.

Figure 4 shows that the results of the qualitative tests clearly supported the quantitative model comparison tests. When the participants knew the cue directions, their estimations were in line with the mapping model's predictions. Similarly, when the participants did not know the cue directions, but all cues were predictive, the participants showed a similar pattern to that predicted by the mapping model. Only in the condition in which the participants did not know the cue directions and only three cues were predictive was the qualitative pattern of the estimations consistent with the exemplar model's predictions.

Discussion of Study 2

Study 2 confirmed the prediction that a rule-based process as described by the mapping model depends on gaining accurate knowledge about the cue directions, which is not necessary for the exemplar model. In the two conditions in which participants were told which cue values were regarded as positive evidence, the mapping model was clearly better in explaining participants' behavior. However, when the participants had to learn the cue directions during training and when this was difficult, because only three cues substantially correlated with the criterion, the exemplar model was the superior model. These results are consistent with the results reported by von Helversen and Rieskamp (2008) and Juslin et al. (2008) and shed light on why the authors had found support for the mapping model in one study but in another the exemplar model was superior.

Although the mapping model clearly outperformed the exemplar model in the generalization test in both conditions in which all cues were predictive, it should be noted that the mapping model predicted the estimations worse when the participants learned the cue directions than when the participants were informed about the directions. This result is partly attributable to some participants who failed to learn the cue directions. The cross-validation test suggests that these participants relied on an exemplar-based approach, while participants who learned the cue directions were better described by the mapping model.

General Discussion

Past research has proposed that multiple distinct processing systems control human cognitive behavior. Which system wins out depends on the structure of the task (e.g., Ashby et al., 1998; Juslin et al., 2008). For instance, explicit, rule-based processes are assumed to be constrained to tasks in which stimulus dimensions are separable and can be selectively attended to, while implicit, similarity-based processes catch on if the stimulus dimensions are integral (Ashby et al., 1998).

Following up on this line of research, our goal was to specify under which conditions two recent models of quantitative estimation, the mapping model (von Helversen & Rieskamp, 2008) and an exemplar model (Juslin et al., 2003, 2008), capture the cognitive processes in quantitative estimations. We derived the exemplar memory prediction and the knowledge abstraction prediction on theoretical considerations of the models' assumptions about the cognitive process to investigate the link between cognitive processing and task characteristics. Assuming that for the exemplar model the accurate storage and retrieval of exemplars influences performance while for the mapping model the establishment of knowledge about the cue directions should be crucial, the knowledge abstraction prediction assumes that the mapping model describes estimations well when knowledge about the task is available or can be easily gained during the task. In contrast, according to the exemplar memory prediction an exemplar-based process might be triggered when the stimulus material allows the accurate storage and retrieval of training exemplars (Ashby & Ell, 2001; Rouder & Ratcliff, 2006). Overall, the results provided clear evidence for task-contingent cognitive processes in accordance with the knowledge abstraction prediction. The mapping model performed best when the participants were informed about the cues' directions or could learn them during training. However, when abstracting knowledge about the cues was difficult but exemplar memory could be used for accurate estimation, the exemplar model was best in predicting participants' estimations. This task contingency between two different cognitive processes could be the result of a learning process reinforcing the reliance on the respective process (Erev & Barron, 2005; Rieskamp, 2006, 2008; Rieskamp & Otto; 2006). However, we did not find evidence to support the exemplar memory prediction. In the following we will discuss the relevance of establishing accurate knowledge abstraction and exemplar memory for quantitative estimations in more detail.

Exemplar Memory: Number of Training Trials and Number of Objects

In categorization research the question of whether the accuracy with which exemplars can be encoded and retrieved influences cognitive processing has been widely discussed. Some research suggests the reliance on rule-based processes or the abstraction of prototypes with large number of exemplars (e.g., Ashby & Ell, 2001; Homa et al., 2008). Similarly, it has been argued that participants might start with testing simple rule-based strategies early in training but later fall back on exemplar-based processing (Johansen & Palmeri, 2002; Karlsson, et al., 2008; Rehder & Hoffman, 2005; Rouder & Ratcliff, 2006). However, other results suggest that the performance of exemplar models is widely independent of the number of training exemplars and the frequency with which exemplars are encountered (e.g., Nosofsky, 1988a, 1988b; Nosofsky & Zaki, 1998).

The results of Study 1 show no effect of the amount of training or the number of training exemplars on cognitive processing. However, these results might be limited to situations with available task knowledge. In Study 1 participants were informed about the cue directions, which apparently strongly influenced the cognitive processing. In sum, our results suggest that in situations in which sufficient knowledge is provided about the task structure people rely on a rule-based estimation process and not on an exemplar-based process even if an exemplar-based strategy could provide a more accurate solution. However, it is not clear if exemplar-based processes might become reinforced by an increased amount of training and a smaller number of training instances, when only little knowledge is available about the task structure.

Prior Knowledge

Providing explicit knowledge about the cue directions led to a strong effect on the estimation process. The mapping model clearly suffered when participants were not informed about the cue directions prior to the task. Furthermore, in the two conditions in Study 2 in which participants had no prior knowledge, the participants performed worse during training, indicating that if knowledge about the task needs to be acquired during training, learning can be impeded (for similar effects of prior knowledge see, e.g., Hoffman, Harris, & Murphy, 2008; Muchinsky & Dudycha, 1975).

However, the exemplar model was only better in predicting participants' estimations when just a subset of the cues substantially correlated with the criterion. This suggests that lacking knowledge about the cue directions was not sufficient to trigger exemplar-based processing, but that the accuracy and difficulty with which rule-based estimation processes could be employed played an important role when exemplar-based processing occurred (see also Ashby et al., 1998; Juslin et al., 2008; Olsson et al., 2006).

The condition with no prior information about the cue directions and only three predictive cues provided especially problematic circumstances for the mapping model, because it affected two of its core assumptions. First, the mapping model requires that accurate knowledge about the cue directions be gained. This was difficult to achieve, as no information about the cues was available and the cue directions were difficult to pick up. Second, the mapping model assumes that all cues are equally important. This suggests that it should be the better model if cues have similar validities. However, in this task, in fact, only three cues were substantially correlated with the criterion. Thus, if participants learned to ignore the less valid cues (Castellan, 1973; Klayman, 1988b), the mapping model should not be able to predict their estimations accurately.

This raises the question of why the mapping model performed well when only three cues were predictive and information about the cue directions was available. The good performance of the mapping model in this condition indicates that participants regarded all cues as equally important for making the estimations, which in turn implies that the participants did not accurately learn the task structure. In fact, to improve their estimation accuracy it would have been advantageous to use only the predictive cues for an estimation. Apparently, providing the participants with explicit knowledge about the direction of the cues led to the inference that all cues were relevant to predict the criterion and thereby triggered a rule-based process consistent with the mapping model.

However, it needs to be noted that we only considered an exemplar model that did not make any use of the prior knowledge concerning the cues' directions. It could be that the estimations were based on an exemplar process that used the available prior knowledge. The categorization literature has illustrated that prior knowledge can affect cognitive processing during learning and categorization (Hoffman et al., 2008; Wisniewski & Medin, 1994). Thus, it is possible that the exemplar model failed in our studies because it did not include a mechanism to represent prior knowledge. However, the good performance of the mapping model even in situations in which prior knowledge about the cue directions was not given, but could easily be acquired, speaks against this conclusion. In contrast, according to the exemplar model it should make no difference how easily the cue directions can be learned. This suggests that the ease with which the models could be applied influenced the cognitive processing. In sum, the conclusions we have drawn about a rule-based versus an exemplar-based processing are limited to the models we have used. In particular, we have applied a standard exemplar model to predict the estimations. More specific exemplar model that make use of prior task knowledge, such as the KRES model for categorization (see Rehder &

Murphy, 2003), could of course provide better accounts of estimations process, but the development and test of such models for estimation problems should be tackled in future research.

Quantitative Model Comparison: Generalization Test Versus Cross-Validation

In this article we relied on two quantitative model-comparison methods: a generalization test and a cross-validation test. The results of both tests mostly supported the same conclusions. However, there were some differences between the tests worth discussing. For one, the exemplar models performed worse in the generalization test than in the cross-validation test. In particular, the standard exemplar model failed in the generalization test, in which it performed worse than the simplified exemplar model. It is possible that the generalization test provides a stronger test of the theories than a cross-validation test: Both tests assume that the cognitive process is the same for the calibration set that is used to estimate the models' parameters and the validation set that is used to test the models' predictions. However, in the case of the generalization test, the calibration and the validation set are not random samples. Therefore in a generalization test the poor accuracy of a model's prediction can be due to a systematic change in cognitive processing between the calibration set and the validation set. In this case the cognitive process could still follow the model's assumptions, but it would reveal a model's weakness in predicting new independent behavior. In other words, the practical value of a model that is not able to predict behavior in a slightly new situation to which it has not been calibrated appears rather small.

Accordingly, there are two explanations for the poor performance of the standard exemplar model in the generalization test. First, the cognitive process could have changed from the training to the test phase, for instance, by giving different attention to the various cues. However, because the training and test sets had a similar structures and the procedure was extremely similar, the assumption of a changed exemplar-based cognitive process appears unlikely. Furthermore, the mapping model fitted participants equally well in the training and test phases, suggesting that no change in processing appeared. Thus, a second more plausible explanation relies on the high flexibility of the exemplar model. The flexibility of the exemplar model could have led to fitting unsystematic error variance for the training phase, which led to poor predictions for the test phase.

Quantitative Estimation and Function Learning

Quantitative estimations share some similarities to function-learning problems (Busemeyer, Byun, Delosh, & McDaniel, 1997; Klayman, 1988b; Slovic & Lichtenstein, 1971). For both types of problems a continuous criterion of a stimulus object has to be predicted. However, while in the quantitative estimation task examined in this article multiple dichotomous cues were available to predict a continuous criterion, most function-learning research focuses on a single continuous cue to predict a continuous criterion. In principle, a version of the mapping model generalized to continuous cues or the exemplar model could be applied to function-learning problems just as well as function-learning models generalized to multidimensional stimuli could be applied to quantitative estimation from multiple cues. Two well-known representatives are the extrapolation-association (EXAM) model (DeLosh, Busemeyer, & McDaniel, 1997) and the population of linear experts (POLE) model (Kalish, Lewandowsky, & Kruschke, 2004). While EXAM combines an exemplar-based process with a rule-based extrapolation mechanism, POLE assumes that a population of “linear experts” is used to represent nonlinear functions. That is, each training object is associated with a specific linear function, the linear expert, which is then used to predict the criterion for this object. So far these two models have only been applied to objects described by one single cue. Nevertheless they could be extended to multiple cues, in which case it would be interesting to test their assumptions against the mapping model and the exemplar model. A possible test situation could be an extrapolation test, as both function-learning models allow for extrapolation processes while the mapping model and the exemplar model do not include an extrapolation mechanism.

Final Conclusion

Previous research has described estimation processes almost exclusively with multiple linear regression models. Recently new cognitively motivated models, such as the exemplar model by Juslin et al. (2008) and the mapping model by von Helversen and Rieskamp (2008; see also Brown & Siegler, 1993) have been proposed to model estimation processes. Interestingly, these models represent two different views on estimation processes. While the exemplar model proposes an implicit, similarity-based process, the mapping model assumes a rule-based process. The experimental studies reported here illustrate the link between the cognitive processes assumed by the models and the structure of the environments. We showed that the models' assumptions about the estimation process were directly affected by different structures of the estimation task, which consequently determined which estimation process prevailed.

References

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

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.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*(5), 204-210.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149-178.

Brehmer, A., & Brehmer, B. (1988).What have we learnt 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: Elsevier/North Holland.

Brehmer, B. (1973). Single-cue probability learning as a function of the sign and magnitude of the correlation between cue and criterion. *Organizational Behavior and Human Decision Processes, 9*, 377-395.

Brehmer, B. (1994). The psychology of linear judgment models. *Acta Psychologica, 87,* 137–154.

Bröder, A., & Gaissmeier, W. (2007). Sequential processing of cues in memory-based multi-attribute decisions. *Psychonomic Bulletin & Review, 14*, 895-900.

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review*, *100*, 511–534.

Brunswik, E. (1952). *Conceptual framework of psychology*. Chicago: University of Chicago Press.

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 K. Lamberts & D. R. Shanks (Eds.), *Knowledge, concepts and categories. Studies in cognition* (pp. 408-437). Cambridge, MA: MIT Press.

Busemeyer, J. R., & Wang, Y-M. (2000). Model comparisons and model selection based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*, 171–189.

Castellan, N. J. (1973). Multiple-cue probability learning with irrelevant cues. *Organizational Behavior and Human Decision Processes, 9*, 16-29.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods and applications*. San Diego: Academic Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571-582.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition, 107*, 284-294.

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, & Cognition, 23*, 968-986.

Doherty, M., & Brehmer, B. (1997). The paramorphic representation of clinical judgment: A thirty-year retrospective. In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on* *judgment and decision making: Currents, connections and controversies* (pp. 537–551). Cambridge: Cambridge University Press.

Ebbesen, E. B., & Konecni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology*, *32*, 805–821.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance, 13*, 171-192.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112*, 912-931.

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

Gigerenzer, G., & Kurz, E. (2001). Vicarious functioning reconsidered: A fast and frugal lens model. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 342-347). New York: Oxford University Press.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 3–34). New York: Oxford University Press.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1-67.

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition, 65*, 197-230.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62,* 255–262.

Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications*. New York: Oxford University Press.

Harries, P. A., & Harries, C. (2001). Studying clinical reasoning. Part 2: Applying social judgment theory. *British Journal of Occupational Therapy, 64*, 285–292.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528-551.

Hoffman, A. B., Harris, H. D., & Murphy, G. L. (2008). Prior knowledge enhances the category dimensionality effect. *Memory & Cognition, 36*, 256-270.

Hoffman, A. B., & Murphy, G. L. (2006). Category dimensionality and feature knowledge: When more features are learned as easily as fewer. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32*, 301-315.

Hogarth, R. M., Gibbs, B. R., McKenzie, C. R. M., & Marquis, M. A. (1991). Learning from feedback: Exactingness and incentives. *Journal of Experimental Psychology: Learning Memory and Cognition, 17*, 734–752.

Hogarth, R. M., & Karelaia, N. (2005). Simple models for multiattribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Management Science, 51*, 1860-1872.

Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning Memory and Cognition, 17*, 444-458.

Homa, D., Proulx, M. J., & Blair, M. (2008). The modulating influence of category size on the classification of exception patterns. *The Quarterly Journal of Experimental Psychology, 61*, 425-443.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology, 45*, 482-553.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition, 106*, 259-298.

Juslin, P., & Olsson, H. (2004). Note on the rationality of rule-based versus exemplar-based processing in human judgment. *Scandinavian Journal of Psychology, 45*, 37-47.

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

Juslin, P., & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): A "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science, 26*, 563-607.

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: Knowledge partitioning and function learning. *Psychological Review, 111*, 1072-1099.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 829-846.

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.

Klayman, J. (1988a). Cue discovery in probabilistic environments: Uncertainty and experimentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 317-330.

Klayman, J. (1988b). On the how and why (not) of learning from outcomes. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 115–162): Amsterdam: North Holland.

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

Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 775-799.

Muchinsky, P. M., & Dudycha, A. L. (1975). Human inference behavior in abstract and meaningful environments. *Organizational Behavior & Human Performance, 17*, 377-391.

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

Nosofsky, R. M. (1988b). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 54-65.

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.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 248-255.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 924-940.

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.

Olsson, H., Wennerholm, P., & Lyxzen, U. (2004). Exemplars, prototypes, and the flexibility of classification models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 936-941.

Patalano, A. L., Smith, E. E., Jonides, J., & Koeppe, R. A. (2001). PET evidence for multiple strategies of categorization. *Cognitive, Affective & Behavioral Neuroscience, 1,* 360-370.

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

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review, 113*, 57-83.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*, 1-41.

Rehder, B., & Murphy, G. L. (2003). A knowledge-resonance (KRES) model of category learning. *Psychonomic Bulletin & Review, 10*, 759-784.

Rieskamp, J. (2006). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1371–1384.

Rieskamp, J. (2008). The importance of learning in probabilistic inferences. *Judgment and Decision Making Journal, 3*, 261-277.

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

Rouder, J. N., & Ratcliff, R. (2006). Comparing exemplar- and rule-based theories of categorization. *Current Directions in Psychological Science, 15*, 9-13.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General, 121*, 278-304.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior & Human Decision Processes, 6*, 649-744.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

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

Wigton, R. S. (1996). Social judgement theory and medical judgement. *Thinking and Reasoning, 2*, 175–190.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18*, 221-281.

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1160-1173.

Zedeck, S., & Kafry, D. (1977). Capturing rater policies for processing evaluation data. *Organizational Behavior and Human Performance, 18*, 269–294.

Appendix A

Selection of Training and Test Sets for Studies 1 and 2

The following tables describe the sets of items that were used in Studies 1 and 2. Table A1 describes the item set for the training phase of Study 1. Table A2 describes the item set for the training and test phases of the condition with a small number of training objects in Study 1 and for the condition with six predictive cues in Study 2. Table A3 describes the set of items for the training and test phases of the conditions with three predictive cues in Study 2.

We first selected the training set–test set combination for the condition with 24 exemplars. To ensure that the training set would well represent the total set, we constrained the selection of training objects to contain objects with all possible cue sums approximately in proportion to the frequency in the whole set. To find a training set–test set combination for which the models made qualitatively different predictions, we generated 100 different training samples and calculated model predictions for the remaining objects based on the respective training sample by fitting the models on the training sample and making predictions based on the estimated parameter values. We selected the training set–test set combination for which the models differed most in their qualitative predictions. The training set consisted of 24 objects and the test set of 22 new objects. Lastly, we included 8 objects in the test set that had appeared in the training set. The training set is reported in Table A1, the test set in Table 1.

To select the training set–test set combination for the condition with eight exemplars we repeated the procedure described above. To make the condition with 8 training exemplars more comparable to the condition with 24 training exemplars, the 100 training sets with 8 training objects were randomly drawn from the condition with 24 training objects, with the restriction that the training sample contained one object each with a cue sum of 0, 1, 2, 4, 5, and 6, respectively, and two objects with a cue sum of 3. Again, we obtained model predictions for the remaining objects and selected a test set that maximized the differences in qualitative predictions. The test set consisted of 22 new objects that were not included in the training set and the 8 known objects from the training phase (see Table A2).

Table A1

Sets of Objects for the Training Phases of Study 1

Training condition	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5	Cue 6	Criterion
A & B	0	0	0	0	0	0	1
A	0	1	0	0	0	0	2
A & B	1	0	0	0	0	0	2
A & B	0	0	0	0	1	1	2
A	0	0	0	1	0	1	3
A	0	1	0	0	0	1	3
A	0	1	0	0	1	0	4
A	1	0	0	0	1	0	4
A	0	0	1	1	1	0	7
A	0	1	0	0	1	1	7
A & B	0	1	0	1	0	1	7
A	0	1	1	0	1	0	9
A	1	0	0	1	0	1	8
A & B	1	0	1	0	1	0	10
A	1	0	1	1	0	0	10
A	1	1	0	0	0	1	10
A	0	1	0	1	1	1	14
A & B	1	1	0	1	1	0	24
A	1	1	1	0	0	1	24
A	1	1	1	0	1	0	26
A	1	1	1	1	0	0	27
A & B	0	1	1	1	1	1	33

Training condition	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5	Cue 6	Criterion
A	1	1	1	1	1	0	55
A & B	1	1	1	1	1	1	100

Note. A & B indicates objects that were used for the training condition (A) with a large number of training objects and for the training condition (B) with a small number of training objects. A indicates objects that were additionally used in the training condition (A) with a large number of training objects.

Table A2

Sets of Objects for the Training and Test Phases of Study 1 for the Condition with a Small Number of Training Objects and of Study 2 for the Condition with Six Predictive Cues

Objects	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5	Cue 6	Criterion	Mapping	Exemplar
Test/training	0	0	0	0	0	0	1	1	1
Test/training	1	0	0	0	0	0	2	2	2
Test/training	0	0	0	0	1	1	2	2	2
Test/training	0	1	0	1	0	1	7	8	7
Test/training	1	0	1	0	1	0	10	8	10
Test/training	1	1	0	1	1	0	24	24	24
Test/training	0	1	1	1	1	1	33	33	33
Test/training	1	1	1	1	1	1	100	100	100
Test 2	0	0	0	1	0	1	3	2	7
Test 2	0	0	0	1	1	0	3	2	9
Test 2	0	0	1	0	1	0	3	2	10
Test 2	0	1	0	0	0	1	3	2	7
Test 2	0	1	0	0	1	0	4	2	9
Test 2	0	1	0	1	0	0	4	2	7
Test 3a	0	0	1	0	1	1	6	8	2
Test 3a	1	1	0	1	0	0	12	8	24
Test 3b	1	0	1	0	0	1	9	8	6
Test 3b	0	1	0	1	1	0	8	8	24
Test 3c	1	0	0	0	1	1	7	8	2
Test 3c	0	1	1	1	0	0	9	8	20
Test 3d	1	0	0	1	0	1	8	8	5
Test 3d	0	1	1	0	1	0	9	8	21
Test 3e	1	1	0	0	0	1	10	8	5
Test 3e	0	0	1	1	1	0	7	8	21
Test 4	1	0	1	0	1	1	17	24	10
Test 4	1	0	1	1	1	0	20	24	10
Test 4	1	1	0	1	0	1	21	24	7
Test 4	1	1	1	0	1	0	26	24	10
Test/extra	1	0	1	1	1	1	37	33	100
Test/extra	1	1	1	1	0	1	50	33	100

Note. Test/training indicates the eight objects that constituted the training set in the condition with a small number of training objects in Study 1 and the two conditions with six predictive cues in Study 2. These eight objects also appeared in the respective test sets. Test 2 denotes objects with a cue sum of 2, Test 3 objects with a cue sum of 3, where pairs with the same letter indicate opposite cue profiles, and Test 4 objects with a cue sum of 4. Test/extra indicates objects that were additionally included in the test set to increase the differences in model predictions.

Table A3

Sets of Objects for the Training and Test Phases of Study 2 for the Condition with Three Predictive Cues

Objects	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5	Cue 6	Criterion	Exemplar	Mapping
Test/training	0	0	1	0	0	0	2	2	2
Test/training	0	0	0	1	0	1	3	3	3
Test/training	0	1	1	0	0	0	4	4	3
Test/training	0	1	1	0	1	0	9	9	8
Test/training	1	0	0	1	0	1	8	8	8
Test/training	1	1	0	1	1	0	24	24	25
Test/training	1	1	1	1	0	0	27	27	25
Test/training	1	0	1	1	1	1	37	37	37
Test 2	0	1	0	0	1	0	4	9	3
Test 2	1	0	0	0	0	1	4	8	3
Test/extra	1	0	0	0	1	0	4	24	3
Test 2	1	0	0	1	0	0	4	8	3
Test/extra	1	1	0	0	0	0	6	18	3
Test 3a	0	0	0	1	1	1	5	3	8
Test 3a	1	1	1	0	0	0	13	16	8
Test 3b	0	0	1	1	0	1	6	3	8
Test 3b	1	1	0	0	1	0	11	24	8
Test 3c	0	1	0	1	0	1	7	3	8
Test 3c	1	0	1	0	1	0	10	16	8
Test 3d	0	1	1	0	0	1	8	4	8
Test 3d	1	0	0	1	1	0	9	24	8
Test 3e	0	1	0	0	1	1	7	9	8
Test 3e	1	0	1	1	0	0	10	27	8
Test/extra	0	1	0	1	1	1	14	13	25
Test 4	0	1	1	0	1	1	16	9	25
Test 4	0	1	1	1	1	0	18	9	25
Test 4	1	1	0	1	0	1	21	8	25
Test 4	1	1	1	0	1	0	26	9	25
Test/extra	1	1	1	1	1	0	55	25	37
Test/extra	1	1	1	1	1	1	100	37	37

Note. Test/training indicates the eight objects that constituted the training set in the two conditions with three predictive cues in Study 2. These eight objects also appeared in the respective test sets. Test 2 denotes objects with a cue sum of 2, Test 3 objects with a cue sum of 3, where pairs with same letters indicate opposite cue profiles, and Test 4 objects with a cue sum of 4. Test/extra indicates objects that were additionally included in the test set to increase the differences in model predictions.

Appendix B

Additional Model Comparison Tests in Studies 1 and 2

For both studies we also tested the performance of a multiple linear regression, a unit weight linear model, and the standard exemplar model with a free parameter for every cue in predicting participants' estimations, in addition to the mapping model and the exemplar model with one attention weight parameter. We conducted a generalization and a cross-validation test.

Generalization test. The predictions of the regression model were obtained by running a multiple linear regression with the cues of the training phase as predictors and participants' estimations in the last four blocks as the dependent variable. On the basis of the obtained cue weights, predictions for the test phase were made. Similarly, the best fitting parameter value for the unit weight linear model was obtained by fitting the model to the estimations of the participants in the training phase. Then with these parameters the predictions for the test phase were generated. Also the exemplar models' parameter values were estimated by using participants' responses in the training phase, and the predictions for the test set were determined based on the estimated parameter value. The best fitting parameter values were calculated with a nonlinear least square method as implemented in MATLAB. For the simplified exemplar model one single attention parameter s was estimated; for the standard exemplar model s_i was allowed to vary freely for each cue i. For the mapping model the model predictions for the test set were calculated based on the cue and criterion values of the training set.

Cross-validation test. For the cross-validation test, the test set was randomly split into two equally sized sets, one for calibration and one for validation. On the basis of the calibration set the parameters for all models were estimated as they were for the generalization test. Then predictions for the validation set were made based on the estimated parameters and the *RMSD* between model prediction and participants' estimates was determined for the validation set. This procedure was repeated 10 times to achieve reliable estimates and the average *RMSD* was used as goodness-of-fit measure. For the exemplar models and the mapping model the training set with its correct criterion values was assumed as a knowledge base.

Study 1

The results of the generalization test are reported in the main body of the article, so we concentrate here on the cross-validation test. The cross-validation test supported the mapping model as the best model to describe participants' estimations (all *Z*s < −3.30, *p*s < .01). Table B1 provides the means and standard deviations of the models' fit. The standard exemplar model performed as well as the simplified exemplar model (Z_{large} = −.82, *p* = .41; Z_{small} = −1.17, *p* = .26) and both exemplar models performed better than the two linear additive models in the condition with a large training set (all *Z*s < −3.92, *p* < .01). In the condition with a small training set both exemplar models performed better than the unit weight regression model (all *Z*s < −2.33, *p* < .02) and as well as the multiple linear regression (*Z*s > −1.61, *p* > .11).

Study 2

In the generalization test, the mapping model was the best model when all cues were predictive, outperforming the regression model, the unit weight model, and the standard and the simplified exemplar models (all *Z*s < −2.80, *p* < .01). The simplified exemplar model performed better than the regression model and the standard exemplar model in both conditions (*Z*s < −2.88, *p* < .01), as well as the unit weight model in the condition with unknown cue directions (*Z* = −.86, *p* = .41), and worse than the unit weight model in the condition with known cue directions ($Z = -3.92$, $p < .01$). In the condition where half of the cues were predictive and cue directions were known, the mapping model again outperformed all other models (all Zs < -3.73, $p < .01$). The unit weight model was second best, outperforming both exemplar models and the regression model (all Zs < -3.29, $p < .01$). In the condition where the cue directions were unknown and only three cues were predictive, it was difficult to determine the best model. The exemplar models, the multiple linear regression, and the unit weight model performed about equally well, with the simplified exemplar model beating the unit weight model ($Z = -3.17$, $p < .01$) but being equally as good as the regression model ($Z = -.30$, $p = .78$) and the standard exemplar model ($Z = -.08$, $p = .96$). The standard exemplar model predicted the estimations of 5 (25%) participants best, the regression model provided the best estimations for 3 (15%), and the simplified exemplar model for 12 (60%). All models performed better than the mapping model (all Zs < -2.24, $p = .02$). An overview of the accuracies of the regression model, the unit weight model, and the standard exemplar model in Study 2 is reported in Table B2.

The pattern of results in the cross-validation test was quite similar. The mapping model was the best model if the cue directions were known (Zs < -2.91, ps $< .01$). The mapping model was also better than the simplified exemplar model and the two linear additive models if all cues were predictive and the cue directions unknown ($Z < -2.05$, $p < .04$) and equally as good as the standard exemplar model ($Z = -.24$, $p = .84$), with 9 participants better predicted by the exemplar model and 11 participants by the mapping model (for means and *SD*s see Table B3). In the condition with only three predictive cues and no prior information the mapping model was outperformed by the exemplar models and the linear additive models (all Zs < -3.62, ps $< .01$). The multiple linear regression model performed better than the unit weight model ($Z = -3.73$, $p < .01$) and the simplified exemplar model (Z = –1.94, p = .05) and equally as well as the standard exemplar model (Z = –.86, p = .41). This suggests that participants in fact weighted the cues differentially.

Table B1

Average Model Accuracies (RMSD) in Predicting the Estimations for the Cross-Validation Test in Study 1

	Number of training objects			
	Large		Small	
	M	*SD*	*M*	*SD*
Exemplar standard	10.07	1.94	13.14	2.42
Regression	14.82	1.41	14.00	2.26
Unit weight	16.06	1.50	14.97	2.54
Mapping	5.69	2.00	6.39	3.92
Exemplar simplified	9.93	1.85	13.95	1.73

Table B2

Average Model Accuracies (RMSD) in Predicting the Estimations of the Regression Model, the Unit Weight Model, and the Standard Exemplar Model for the Generalization Test in Study 2

	Number of training objects							
	Six predictive cues				Three predictive cues			
	Cue directions				Cue directions			
	Known		Unknown		Known		Unknown	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Training set								
Exemplar standard	1.20	1.72	7.31	6.96	2.45	1.69	4.53	2.58
Regression	12.89	0.30	14.70	2.54	2.92	1.04	4.24	1.69
Unit weight	20.98	0.46	21.86	1.68	7.60	0.27	8.23	0.98
Mapping	1.77	1.77	8.86	8.56	2.89	1.75	6.03	3.94
Exemplar simplified	1.40	2.03	8.40	8.07	2.68	1.78	5.61	3.63
Test set: Old objects								
Exemplar standard	3.19	3.82	8.35	6.30	3.07	2.35	5.35	2.49
Regression	13.33	0.92	15.61	3.04	3.56	1.90	5.63	2.72
Unit weight	20.99	1.03	21.99	2.42	7.66	0.71	8.34	1.07
Mapping	3.44	3.73	8.39	7.94	3.12	2.46	6.35	3.22
Exemplar simplified	3.25	3.90	8.92	8.05	2.99	2.44	6.02	3.03
Test set: New objects								
Exemplar standard	29.94	8.20	28.18	6.68	16.57	3.30	10.05	3.10
Regression	27.03	2.57	27.94	3.52	13.90	3.57	10.01	2.54
Unit weight	18.00	2.02	21.91	3.83	12.96	3.70	10.15	1.16
Mapping	6.34	4.00	16.34	22.22	10.36	4.50	12.24	2.21
Exemplar simplified	23.50	2.85	7.36	4.29	14.78	3.47	8.71	1.92

Test set: All objects								
Exemplar standard	25.75	7.10	24.72	5.72	14.31	2.88	9.17	2.49
Regression	24.16	2.15	25.28	3.10	12.08	3.06	9.14	2.22
Unit weight	18.86	1.52	21.99	3.05	11.84	2.97	9.71	1.04
Mapping	5.98	3.47	14.88	7.02	9.11	3.86	11.09	1.95
Exemplar simplified	20.29	2.48	19.90	4.20	12.80	3.01	8.16	1.94

Note. N = 80, with n = 20 in each condition.

Table B3

Average Model Accuracies (RMSD) in Predicting the Estimations for the Cross-Validation Test in Study 2

	Number of predictive cues							
	Six predictive cues				Three predictive cues			
	Cue directions				Cue directions			
	Known		Unknown		Known		Unknown	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Exemplar standard	13.34	2.37	15.00	5.43	11.97	2.78	7.35	2.60
Regression	14.44	2.27	18.74	4.41	9.51	3.31	7.14	2.02
Unit weight	15.24	2.23	20.18	4.20	12.16	3.06	9.58	1.09
Mapping	5.35	2.72	14.91	7.09	8.60	3.65	11.08	2.05
Exemplar simplified	14.29	2.19	17.57	5.22	12.57	3.21	7.99	1.76

Note. $N = 80$, with $n = 20$ in each condition.

Footnotes

1. It should be noted that the mapping model and a unit weight linear model make equivalent predictions if the medians of the categories are equidistant.

2. Additionally, the feedback algorithm incorporated a correction term that determined the deviation that would result in a payoff of zero. It was calculated on the basis of a baseline model that always estimated the average criterion value. Any deviation exceeding the correction term led to the subtraction of points. To exclude the subtraction of a high number of points due to a typing error, the feedback algorithm was truncated. Any deviation larger than 50 was treated as a deviation of 50. A similar feedback algorithm had been successfully used by von Helversen and Rieskamp (2008) to create a moderately exacting feedback environment (Hogarth, Gibbs, McKenzie, & Marquis, 1991).

3. We focused on the predictions of the simplified exemplar model with only one parameter because previous studies have indicated that the full version of the exemplar model is prone to over-fitting and it performed worse in the generalization test.

4. To prevent participants from becoming discouraged by overly negative feedback in the beginning of the study, we truncated the feedback algorithm, similar to in Study 1. However, to counteract the higher difficulty in the conditions with no prior information, we decreased the maximum deviation: In Study 2 any deviation larger than 30 was treated as a deviation of 30.

5. We also tested a version of the mapping model that included only the three cues that were substantially correlated with the criterion. However, overall this model did not perform better than a mapping model that considered all cues.

Authors' Note

Bettina von Helversen, Max Planck Institute for Human Development, Berlin, Germany; Jörg Rieskamp, University of Basel, Department of Psychology, Switzerland. We would like to thank Anita Todd for editing a draft of this manuscript. This work has been supported by a doctoral fellowship of the International Max Planck Research School LIFE to the first author and a research grant by the German research foundation to both authors (RI 1226/5). Correspondence concerning this article should be addressed to Bettina von Helversen.

Bettina von Helversen

Max Planck Institute for Human Development

Königin-Luise-Str. 5, 14195 Berlin, Germany

Phone: ++49-30-82406699

Fax: ++49-30-82406394

Email: vhelvers@mpib-berlin.mpg.de

Tables

Table 1

New Test Objects in the Condition with a Large Number of Training Objects of Study 1

Objects	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5	Cue 6	Criterion	Mapping	Exemplar
Test 2	0	0	0	1	1	0	3	3	7
Test 2	0	0	1	0	1	0	3	3	8
Test 2	0	0	1	1	0	0	3	3	8
Test 2	1	0	1	0	0	0	5	3	7
Test 4	1	0	0	1	1	1	16	24	8
Test 4	1	0	1	1	0	1	18	24	9
Test 4	1	1	0	0	1	1	20	24	8
Test 4	1	1	0	1	0	1	21	24	8
Test 3a	0	0	0	1	1	1	5	8	6
Test 3a	1	1	1	0	0	0	13	8	26
Test 3b	0	0	1	0	1	1	6	8	2
Test 3b	1	1	0	1	0	0	12	8	25
Test 3c	0	0	1	1	0	1	6	8	3
Test 3c	1	1	0	0	1	0	11	8	14
Test 3d	0	1	0	1	1	0	8	8	14
Test 3d	1	0	1	0	0	1	9	8	24
Test 3e	1	0	0	0	1	1	7	8	3
Test 3e	0	1	1	1	0	0	9	8	27
Test/extra	1	0	1	0	1	1	17	24	10
Test/extra	0	1	1	0	1	1	16	24	16
Test/extra	0	0	0	0	1	0	1	2	3
Test/extra	1	0	1	1	1	1	37	44	100

Note. Test 2 denotes objects with a cue sum of 2, Test 3 objects with a cue sum of 3, where pairs with the same letter indicate opposite cue profiles, and Test 4 objects with a cue sum of 4. Test/extra indicates objects that were additionally included in the test set to increase the differences in model predictions.

Table 2

Model Accuracies in Predicting Estimations for Study 1

	Number of training objects			
	Large		Small	
	M	*SD*	*M*	*SD*
Training set				
Exemplar standard	4.14	1.45	3.21	2.51
Regression	13.06	0.54	15.70	0.66
Unit weight	16.05	0.53	20.91	0.70
Mapping	5.37	1.12	3.63	2.77
Exemplar simplified	4.38	1.66	3.53	2.81
Test set: Old objects				
Exemplar standard	3.46	2.76	6.31	7.90
Regression	19.40	0.76	16.93	3.09
Unit weight	22.87	0.72	21.13	2.13
Mapping	5.28	1.70	5.94	8.24
Exemplar simplified	3.27	2.54	5.86	8.21
Test set: New objects				
Exemplar standard	16.89	4.86	34.76	7.99
Regression	8.96	1.57	27.86	3.65
Unit weight	11.03	1.34	17.83	1.64
Mapping	5.88	2.32	5.74	3.52
Exemplar simplified	15.45	2.37	22.63	1.82
Test set: All objects				
Exemplar standard	14.65	2.76	30.23	6.74
Regression	12.66	0.97	25.47	3.05
Unit weight	15.14	0.87	18.81	1.31
Mapping	5.80	1.93	6.63	4.02
Exemplar simplified	13.39	2.10	20.00	2.06

Note. $N = 39$, with $n = 20$ in the condition with a large training set and $n = 19$ in the condition with a small training set.

Table 3

Cue–Criterion Correlations in Study 2

	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5	Cue 6
Criterion (six predictive cues)	0.37	0.60	0.63	0.60	0.47	0.43
Criterion (three predictive cues)	0.79	0.15	0.17	0.58	0.56	0.11

Table 4

Model Accuracies in Predicting Estimations for Study 2

	Number of predictive cues							
	Six predictive cues				Three predictive cues			
	Cue directions				Cue directions			
	Known		Unknown		Known		Unknown	
	Mapping	Exemplar	Mapping	Exemplar	Mapping	Exemplar	Mapping	Exemplar
Training set								
RMSD	1.77	1.40	8.86	8.40	2.89	2.68	6.03	5.61
SD_{RMSD}	1.77	2.03	8.56	8.07	1.75	1.78	3.94	3.63
Test set: Old objects								
RMSD	3.44	3.25	8.39	8.92	3.12	2.99	6.35	6.02
SD_{RMSD}	3.73	3.90	7.94	8.05	2.46	2.44	3.22	3.03
Test set: New objects								
RMSD	6.34	23.50	16.34	22.22	10.36	14.78	12.24	8.71
SD_{RMSD}	4.00	2.85	7.36	4.29	4.50	3.47	2.21	1.92
Test set: All objects								
RMSD	5.98	20.29	14.88	19.90	9.11	12.80	11.09	8.16
SD_{RMSD}	3.47	2.48	7.02	4.20	3.86	3.01	1.95	1.94

Note. $N = 80$, with $n = 20$ in each condition. *RMSD* is root mean squared deviation

Figure Captions:

Figure 1.

Qualitative model predictions. The models' predictions for the two qualitative tests, when varying the values of the exemplar model's attention parameter *s*. The "4 vs. 2" denotes the predicted average differences in estimations for the criterion values of test objects with a cue sum of 4 and test objects with a cue sum of 2. The "3" refers to the predicted average differences in estimations for the criterion values of the pair of test objects with a cue sum of 3, with maximally different cue profiles (e.g., 111000 and 000111). (A) The predictions for the condition with a small number of training objects. (B) The predictions for the condition with a large number of training objects.

Figure 2.

Qualitative model comparison test in Study 1. (A) Qualitative predictions of the models and the participants' estimations in the condition with a small number of training objects (n = 19). (B) Qualitative predictions of the models and the participants' estimations in the condition with a large number of training objects (n = 20). Sum of cue values "3" gives the average difference in estimations for the criterion values of the pair of test objects with a cue sum of 3 with maximally different cue profiles. Sum of cue values "4 vs. 2" gives the average difference in estimations for the criterion values of test objects with a cue sum of 4 and test objects with a cue sum of 2; error bars denote ±1 *SD*.

Figure 3.

Models' accuracy in predicting the participants' estimations for the new objects in the test phase of Study 2. (A) Models' accuracy when the cues' directions were known (N = 40, 20 for each condition). (B) Models' accuracy when the cues' directions were not known (N = 40, 20 in each condition).

Figure 4.

Qualitative test in Study 2. (A) Qualitative tests for the condition with known cue directions and three predictive cues. (B) Qualitative tests for the condition with known cue directions and six predictive cues. (C) Qualitative tests for the condition with unknown cue directions and three predictive cues. (D) Qualitative tests for the condition with unknown cue directions and six predictive cues. Sum of cue values "3" gives the average difference in estimations for the criterion values of the pairs of test objects with a cue sum of 3 with maximally different cue profiles. Sum of cue values "4 vs. 2" gives the average difference in estimations for the criterion values of test objects with a cue sum of 4 and test objects with a cue sum of 2. Error bars denote ±1 *SD*; n = 20 in each panel.

Figures

Figure 1.

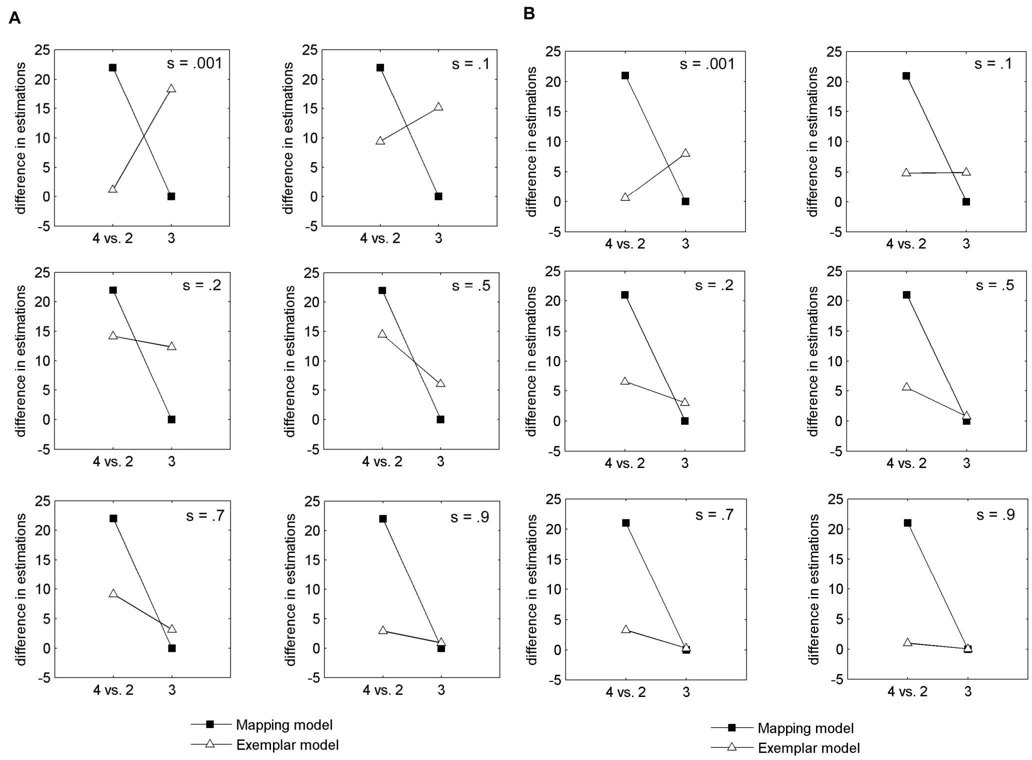


Figure 2.

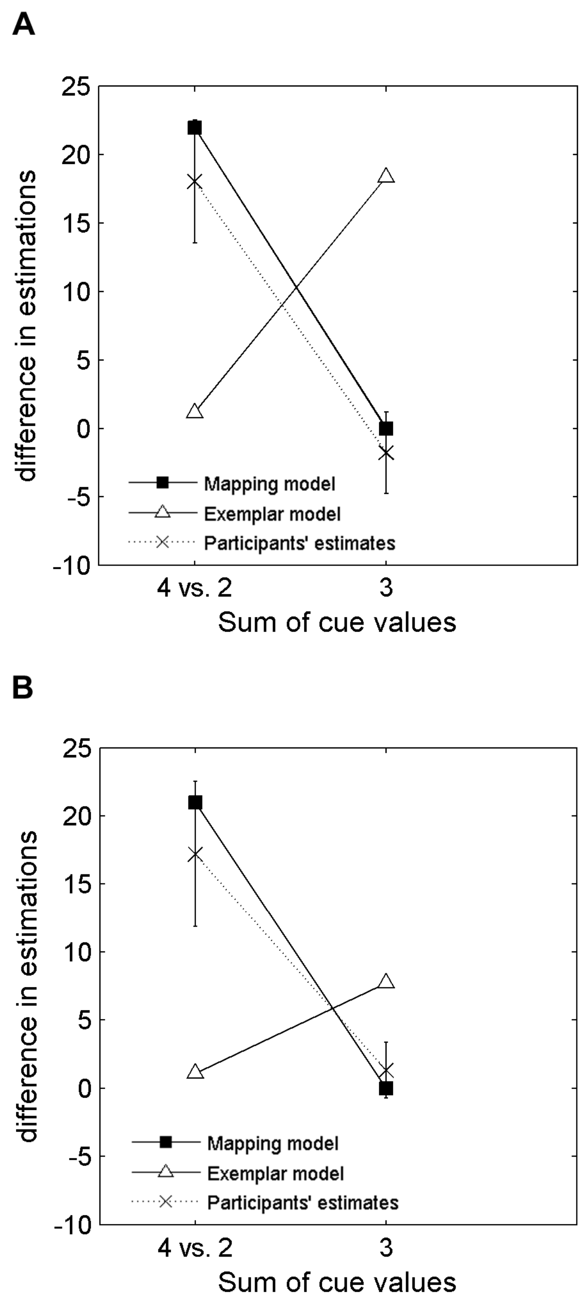

A
25
20
15
10
5
0
-5
-10
difference in estimations
Mapping model
Exemplar model
Participants' estimates
4 vs. 2
3
Sum of cue values
B
25
20
15
10
5
0
-5
-10
difference in estimations
Mapping model
Exemplar model
Participants' estimates
4 vs. 2
3
Sum of cue values

Figure 3.

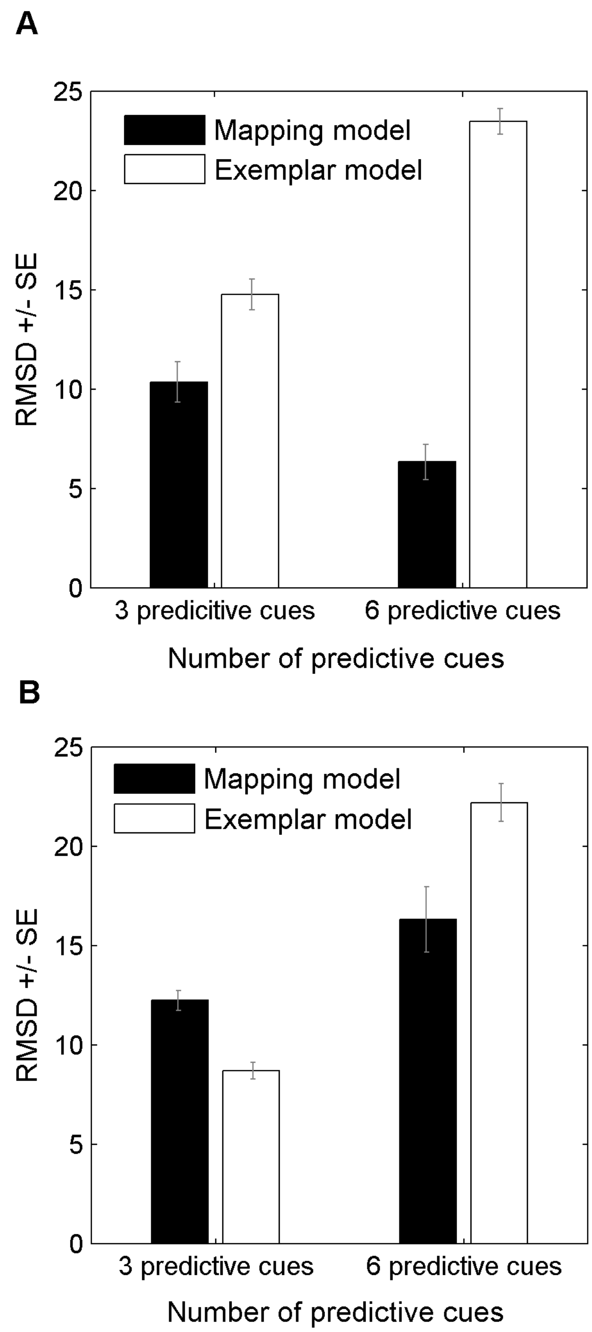
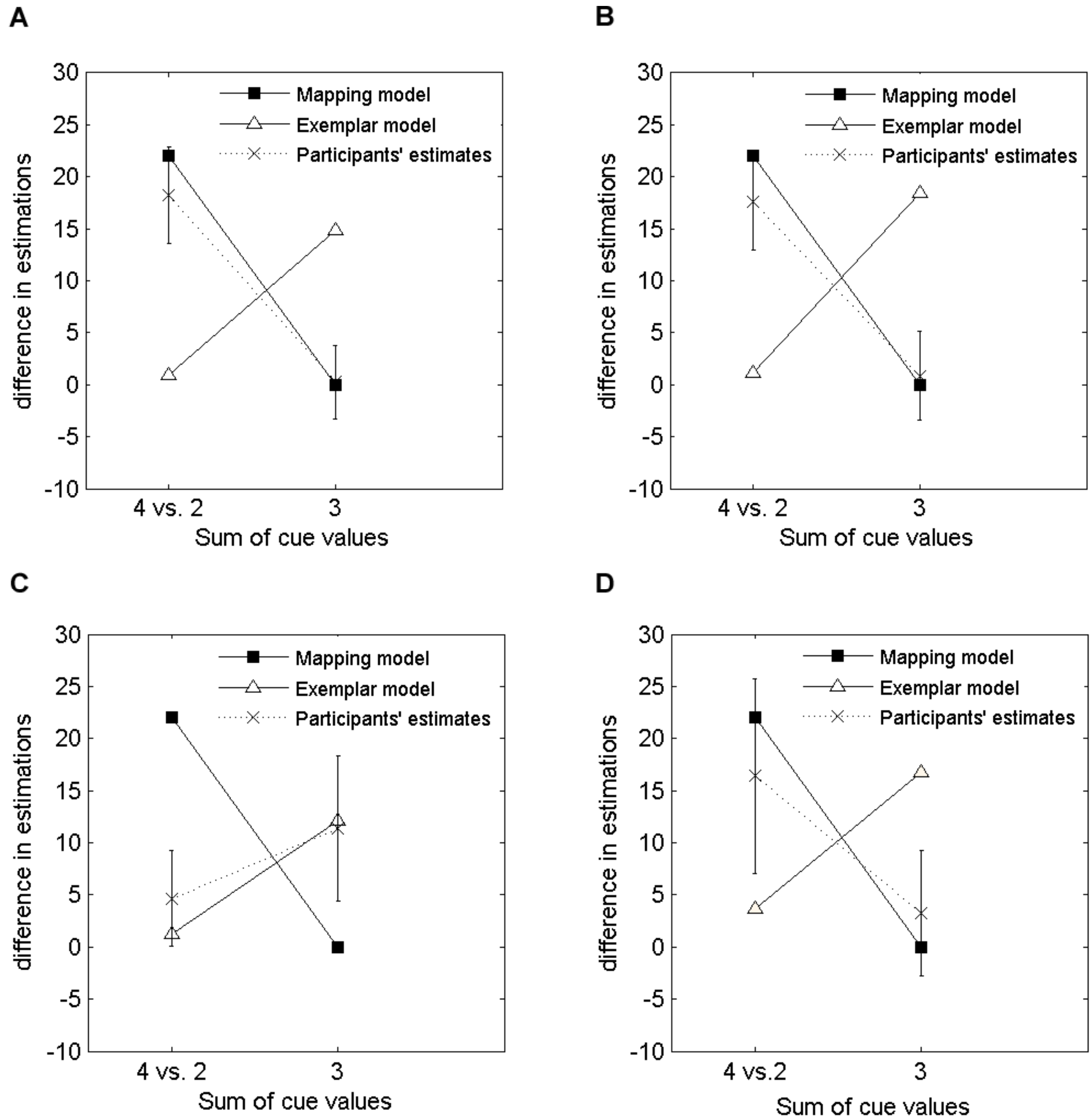

A
Mapping model
Exemplar model
RMSD +/- SE
0
5
10
15
20
25
3 predicitive cues
6 predictive cues
Number of predictive cues
B
Mapping model
Exemplar model
RMSD +/- SE
0
5
10
15
20
25
3 predictive cues
6 predictive cues
Number of predictive cues

Figure 4


A
B
C
D
Mapping model
Exemplar model
Participants' estimates
difference in estimations
Sum of cue values
4 vs. 2
4 vs.2
3
30
25
20
15
10
5
0
-5
-10