Acta Psychologica 130 (2009) 25–37

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Inferences from memory: Strategy- and exemplar-based judgment models compared

Magnus Persson [a,*], Jörg Rieskamp [a,b]

[a] Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
[b] University of Basel, Department of Psychology, Missionsstrasse 62a, 4055 Basel, Switzerland

ARTICLE INFO

Article history:
Received 31 August 2007
Received in revised form 17 September 2008
Accepted 21 September 2008
Available online 4 November 2008

PsycINFO classification:
2340

Keywords:
Decision making
Heuristics
Inference
Memory

ABSTRACT

What are the cognitive processes underlying people's inferences from memory? To provide an answer, the exemplar-based approach to predicting people's inferences is tested against the strategy-based approach. Exemplar models assume that people make inferences about objects by retrieving similar objects from memory. In contrast, the strategy-based approach assumes that people select cognitive strategies that make inferences based on abstracted knowledge and information the inference situation provides. In Experiment 1, in which dichotomous feedback on the level of pair-comparisons was provided, almost all participants were classified as using a simple lexicographic strategy. In Experiment 2, in which continuous feedback for single objects was provided, most participants were classified as using a compensatory strategy. Both experiments suggest that the strategy-based approach is more suitable for predicting people's inferences from memory than the exemplar-based approach. The strategy-based approach shows how people adapt to inference situations by selecting different cognitive strategies.

1. Introduction

Inferences are often made by retrieving information from memory. When, for instance, inferring which of two meals at a restaurant has fewer calories, the inference is often based on information taken from several cues, some of which are retrieved from memory. How do people make such everyday inferences? Surprisingly, the existing literature focuses more on situations where inferences are made from given information that does not need to be retrieved from memory. In contrast, the present article focuses on inferences that are based on memory and examines how these inferences can be best described.

Decisions between alternatives are often preceded by an evaluation of the available choice alternatives (Einhorn & Hogarth, 1981). These evaluations could be either independent judgments for each alternative separately, or relative judgments comparing alternatives. Considering the vast amount of research on judgments about single alternatives (e.g., for an overview see Cooksey (1996), Dhami, Hertwig, and Hoffrage (2004), and Hammond (1996)), the research on judgments comparing alternatives appears rather limited. This is surprising considering the essential role of relative judgments for decision making. Thus, we consider a pair comparison task in which people infer which of two alternatives has a higher criterion value on the basis of several cues. For example, a decision between two travel routes can be preceded by an inference about which route will take more time. This inference could be made by using cues such as the routes' lengths, the time of the day, or the routes' number of traffic lights.

Recently the study of inferences comparing alternatives has gained increasing attention (Bergert & Nosofsky, 2007; Bröder, 2000, 2003; Bröder & Schiffer, 2003; Dieckmann & Rieskamp, 2007; Garcia-Retamero & Rieskamp, in press; Lee & Cummins, 2004; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003; Nosofsky & Bergert, 2007; Rakow, Hinvest, Jackson, & Palmer, 2004; Rieskamp, 2006a, 2008; Rieskamp & Otto, 2006). This work relies on the core assumption that the cognitive processes underlying inferences can be best described by strategies or rules that make use of the available cues for an inference. We call this the *strategy-based approach* to predicting people's inferences. The strategy-based approach assumes that people have abstract knowledge about the inference situation, such as how cues are associated with the criterion, which they use to make an inference. An alternative perspective called the *exemplar-based approach* can be found in Juslin and Persson (2002) (see also Juslin, Jones, Olsson, and Winman (2003), Juslin, Olsson, and Olsson (2003), and Nilsson, Olsson, and Juslin (2005)). According to the exemplar-based approach inferences are made by comparing the present exemplar

* Corresponding author. Tel.: +49 30 824060; fax: +49 30 8249939/06394.
E-mail addresses: persson@mpib-berlin.mpg.de, persson@phmp.se (M. Persson).

0001-6918/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2008.09.010

(probe) with similar exemplars in memory and by making an inference based on the information provided by the memorized exemplars. Both approaches can be applied to predict people's inferences. For both it is beneficial if the individual has acquired some experience with the inference situation and is provided with feedback about correct inferences. According to the strategy-based approach, the experience allows the decision maker to abstract knowledge that the various strategies require. According to the exemplar-based approach, the experience leads to memories of past exemplars. How people gain experience with an inference situation might also determine how they make inferences and consequently which of the two approaches predicts the inferences best.

In this article, we will make the novel contribution of comparing the two approaches against each other for predicting inferences from memory. First we will describe the two approaches in detail before elaborating on how we compare them. We will then report the results of two experiments that tested the approaches against each other. Finally, we will discuss the results in terms of which of the two approaches is more suitable for predicting people's inferences.

1.1. The strategy-based approach to inferences

Various researchers have argued that people select strategies for decision making (e.g., Ginossar & Trope, 1987; Payne, Bettman, & Johnson, 1993; Rapoport & Wallsten, 1972; Svenson, 1979). For instance, Gigerenzer and Goldstein (1996) have argued that people use simple heuristics when making inferences. The interest in heuristics has been fueled by the counterintuitive claim that ignoring information is often better than considering and integrating all available information. This claim has inspired several experimental studies (e.g., Bergert & Nosofsky, 2007; Bröder, 2000; Dieckmann & Rieskamp, 2007; Lee & Cummins, 2004; Newell & Shanks, 2003; Newell et al., 2003; Nosofsky & Bergert, 2007; Rieskamp, 2006a; Rieskamp & Hoffrage, 1999, 2008) showing that people do select simple heuristics when high application costs exist (e.g., cognitive costs, information search costs, and costs of time pressure) or when the heuristics outperform more complex strategies. However, the experimental evidence also showed that people often select compensatory strategies that integrate a substantial amount of information, for instance, when application costs are low. Thus, people's cognitive processes and the strategies they select often depend on the structure of the decision environment (Payne et al., 1993; Rieskamp, 2006b; von Helversen & Rieskamp, 2008).

1.2. The exemplar-based approach to inferences

Juslin and Persson (2002) argued for a radically different approach to inferences. When making an inference about alternatives, according to the exemplar model the alternatives' cue values are used to retrieve *similar* instances from memory. The inference is then based upon whatever knowledge the retrieved instances provide, rather than on the information gleaned directly from the cues. Exemplar models can also be very accurate, especially when provided with a large memory capacity (Hastie, Tibshirani, & Friedman, 2001). In particular, exemplar models can perform well in situations where the cues are not linearly related to the criterion, where strategies that rely on linear relationships fail (Juslin, Karlsson, & Olsson, 2008; Nosofsky & Palmeri, 1998). Juslin and Persson also showed that exemplar models can be as accurate as the heuristics considered by Gigerenzer and Goldstein (1996) by also using only very limited information.

In general, exemplar models are successful theories of perceptual categorization (for a good overview see Johansen and Palmeri (2002)). However, it is an open question to what extent exemplar models also provide a good account of people's probabilistic inferences where declarative knowledge has to be processed. Recently, Juslin and colleagues (Juslin, Jones, et al., 2003; Juslin, Olsson, et al., 2003) showed that exemplar models can predict inferences well, in particular when it is difficult to abstract the relationships between the cues and the criterion. In most of the experimental studies reported above that provide evidence for the strategy approach, these relationships, specified as cue validities, were provided to the participants. In contrast, Nilsson et al. (2005) did not provide cue validities to their participants and showed that an exemplar model did well in predicting probability judgments. Thus, if information about the relationships between the cues and the criterion is not provided, exemplar models might be more appropriate models of people's inferences.

The empirical support for exemplar models has been limited to inferences about single objects, rather than inferences about two or more alternatives. Juslin, Jones, et al. (2003) included a pair-comparison test in their experiment, but the training with feedback was done using single objects. In this situation the exemplar model worked well in predicting the inferences. However, as Rieskamp and Otto (2006) have pointed out, because in the training phase only single objects were provided, it is not surprising that participants stuck to the exemplar-based inference process and did not switch to a strategy-based process for pair comparisons in the test phase. In sum, despite the growing experimental evidence of how people make inferences, the strategy-based approach has not been satisfyingly compared with the exemplar-based approach. Whereas Rieskamp and Otto's (2006) empirical results support the strategy approach for situations in which people had easy access to the cues (i.e., by using a computerized information board), Juslin, Olsson, et al. (2003) found support for the exemplar approach. Given this conflicting evidence it should be fruitful to examine inferences from memory in more detail and to examine the characteristics of inference situations that trigger a strategy-based or an exemplar-based inference process.

1.3. Inferences from memory versus inferences from givens

Gigerenzer and Goldstein (1996) distinguished inferences with *search from memory* from inferences with *search from givens*, the first referring to situations in which the information for an inference has to be retrieved from memory and the second to the situations in which the information is given, for example, by using an information board in a laboratory setting. They argued that people are more likely to select a simple heuristic called take the best (TTB) when they search for information from memory. TTB searches for the cue with the highest validity. When the cue discriminates, that is, when one alternative has a positive and the other a negative (or missing) cue value, the alternative with the positive cue is selected. If the cue does not discriminate then the second most valid cue is considered, and so on. If no cue discriminates, TTB selects randomly. The validity of a cue is defined as the conditional probability of making a correct inference on the condition that the cue discriminates. Ironically most experiments testing TTB have typically examined inferences with search from givens (e.g., Bröder, 2000; Lee & Cummins, 2004; Newell & Shanks, 2003; Rieskamp, 2006a). Recently it has been discussed to what extent TTB represents a plausible model for people's inferences from memory (Dougherty, Franco-Watkins, & Thomas, 2008; Gigerenzer, Hoffrage, & Goldstein, 2008). This discussion reveals the need for further experimental tests of models for inferences from memory, which we provide with the present research.

The challenge to test TTB under conditions with search from memory was taken up by Bröder and Schiffer (2003). In their elaborate experiments, participants first faced a memorization phase in which they had to memorize information about murder suspects. Thereafter, they had to infer which suspect most likely committed a crime. The results indeed confirmed the "search from memory" hypothesis: The majority (64%) of participants was classified as selecting TTB when they had to retrieve information from memory, whereas only 32% were classified as selecting TTB in the control group with search from givens, where information was conveniently presented with a computerized information board. In the experiment reported here, we made use of Bröder and Schiffer's experimental paradigm to compare the strategy-based approach with the exemplar-based approach. In contrast to Bröder and Schiffer we did not give participants information about the cue–criterion relationships, which could have fostered the application of inference strategies that require this information. Instead, the relationship between the cues and the criterion had to be learned. Bröder and Schiffer did not test an exemplar model for predicting participants' inferences. Therefore, it is an open question whether exemplar models, as opposed to inference strategies, might be more suitable for predicting inferences from memory.

1.4. Two approaches to inferences

In the following we will specify the strategy- and exemplar-based approaches in detail and describe how the two approaches can be compared against each other.

1.4.1. Strategy-based inferences

We focus on an inference task where the decision maker has to infer which of two alternatives has a higher criterion value on the basis of four cues. For such an inference problem various inference strategies can be applied. Rieskamp and Otto (2006) proposed a learning theory that specifies how people select strategies from their strategy repertoire. However, instead of considering this general theory of strategy selection, we will focus on the building blocks of the theory, that is, the single strategies that can be applied to making inferences. The first strategy we address is TTB, described above. To illustrate the application of TTB consider the inference situation in which one alternative has a positive cue value for the most valid cue and negative (respectively, non-positive) cue values for three other cues (represented by a "cue profile" of 1, 0, 0, 0). TTB infers that this alternative has a higher criterion value than another alternative where the most valid cue has a negative cue value and all remaining cues have positive values (providing the profile 0, 1, 1, 1), because the information of the first cue is not compensated for by all the other cues.

As an alternative to TTB we consider a compensatory strategy (COMP) that multiplies the cue values with their weights for each alternative, determines the sum of the weighed cue values, and selects the alternative with the largest sum. In a situation where the cues' validities are provided it is natural to assume that the validities are used as weights. However, in a situation where no validities are provided people need to form their own subjective weights. Thus, COMP represents a class of compensatory strategies defined by the space of possible weights. In the extreme case, when using noncompensatory weights such that a cue cannot be outweighed by any combination of less valid cues, COMP could make identical predictions to TTB (for details see Martignon and Hoffrage (2002)). Therefore, to discriminate between COMP and TTB, we restrict the weights of COMP to plausible compensatory ones, so that a cue can be outweighed by a combination of less valid cues. We will also distinguish a special case of COMP using unit weights, which we will call TALLY.

TALLY counts the number of positive cues for each alternative and selects the alternative with the larger sum. In cases where the sum of positive cue values is the same for both alternatives, TALLY is defined as predicting a random choice (see also Bröder and Schiffer (2003)). Nevertheless, for cases in which TALLY does not discriminate we think it could also be possible that people who apply TALLY will try to discriminate the alternatives, presumably by giving larger weight to some cues. Naturally there are many other strategies people could apply when making inferences. However, those we have selected cover a sufficient range of strategies: TTB represents strategies that focus on single pieces of information and do not rely on trade-offs, COMP represents strategies that integrate a lot of information and make trade-offs, and finally TALLY represents a compensatory strategy that is simple to apply. Thus, we think with this selection of strategies we will be able to predict people's inferences well if they rely on inference strategies. Even if a person had used a different strategy its predictions would be highly correlated with the predictions of one of the strategies we selected. Therefore, someone using an inference strategy would most likely be assigned to using one of the proposed strategies. Our selection of strategies is also supported by the empirical evidence described above, which shows that the selected strategies often predict people's inferences well under various conditions (e.g., Rieskamp & Hoffrage, 2008). Examining which of the various strategies does best in predicting people's inferences from memory is in itself interesting: When information has to be retrieved from memory this requires substantial cognitive effort, making it likely that people more frequently use simple heuristics such as TTB that require little information (see also Bröder and Schiffer (2003)). The two experiments reported below test this prediction.

1.4.2. Exemplar-based inferences

The context model proposed by Medin and Schaffer (1978) has been developed into a large class of models used mainly in categorization research (Kruschke, 1992; Lamberts, 2000; Nosofsky, 1986) and are now generally referred to as exemplar models. There are two principles that are common for all exemplar models: First, numerous distinct instances of stimuli are represented in memory, and second, the similarity of a presented stimulus to old instances is the mechanism behind the retrieval process that ultimately determines the response (Nosofsky, 1992). Algorithms implementing the idea of exemplar models are frequently used in artificial intelligence and are referred to as lazy algorithms because they do not require pre-calculations before the task is encountered (Aha, 1997).

PROBEX (Juslin & Persson, 2002) is an exemplar model that estimates the unknown criterion of an alternative on the basis of retrieved exemplars from memory. Here we use a simplified version of PROBEX. The alternative's criterion is estimated by taking a weighted average of the exemplar's criterion values, where each criterion value is weighted by the similarity of the exemplar and the target alternative. The similarity $S(\bar{o}, \bar{y})$, between an option $\bar{o}$ and an exemplar in memory $\bar{y}$ (using bars for vector notation) is defined by the multiplicative rule:

$$S(\bar{o}, \bar{y}) = s^d, \quad 0 \leqslant s \leqslant 1, \tag{1}$$

where d is the number of cues that differ between $\bar{o}$ and $\bar{y}$. The similarity parameter s governs how sensitive the similarity measure is to d. When the alternative and the exemplar are identical, the similarity equals 1, whereas the similarity asymptotically approaches 0 as a function of the number of different cue values d. To simplify for the purpose of the present article we assume that all exemplars in memory are retrieved (which is in line with traditional exemplar models for perceptual categorization; see, for example, Nosofsky (1986, 1992)). This is one aspect in which our exemplar model differs from PROBEX, where the retrieval process terminates when the difference between two consecutive estimates is less than a threshold.[1] The criterion value of an option is estimated by

[1] The original PROBEX exemplar model assumes that each exemplar is retrieved with a probability proportional to the similarity between the exemplar and the given target alternative. It is an open question how the similarity between the exemplar and the given target alternative is determined.

$$\hat{c}(\bar{o}) = \frac{\sum_{i=1}^{N} S(\bar{o}, \bar{x}_i) \cdot c(\bar{x}_i)}{\sum_{i=1}^{N} S(\bar{o}, \bar{x}_i)}, \quad (2)$$

where N is the total number of exemplars in memory and $c(\bar{x}_i)$ is the criterion associated with exemplar i. This simplified version of PROBEX has only one free parameter, namely, the similarity parameter s, which for the sake of parsimony we set to a fixed default value of $s = 0.5$ (for details see Juslin and Persson (2002)).[2]

1.5. Predictions of the models and their experimental tests

Although the strategy and exemplar approaches are conceptually very different, surprisingly, they often predict the identical choices. This is due to the nature of a pair comparison task. The different models often do lead to different evaluations of the two alternatives that are compared; however, because these evaluations are transformed to a dichotomous decision, any differences in evaluations with respect to the two alternatives are frequently wiped out, making it an experimentally demanding task to test the models against each other. We aimed for an experimental design in which (a) the models would make different predictions, (b) the cue validities would be substantially different so that the participants could distinguish among them, and (c) the inference problem could be solved equally well by the different approaches (to prevent people from learning the more accurate approach). We planned an experiment with three phases: in the first phase (memorization) participants had to memorize the cue values of a set of objects. In the second phase (training) they had to make pair comparisons for a subset of the objects when outcome feedback was provided. Finally, in the third phase (test) the pair comparisons were made for the remaining objects without feedback. This experimental procedure shares many similarities to the approach taken by Bröder and Schiffer (2003).

To keep the memorization phase manageable, we decided to describe each object with four cues, which results in 16 profiles of possible combinations of cues. From these profiles 13 were used for the experiment, 6 for the training phase, and 7 for the test phase (see Tables 1 and 2). For each profile of the training phase the criterion value was computed as the sum of the cue values multiplied by the weights 8, 4, 2, and 1 for the cues A, B, C, and D, respectively, with the exception of alternative P5, to which the criterion value 16 was assigned. These profiles were chosen so that the strategies and the exemplar model would make different predictions in the test phase of the experiment.

Table 1
The cue combinations used for the training phase in the experiments and the criterion used to construct the dichotomous feedback in Experiment 1 or explicit continuous feedback in Experiment 2

	Cues (symptoms)				Criterion value
	A	B	C	D	
Cue profile					
P1	0	0	0	0	0
P2	0	0	0	1	1
P3	0	0	1	0	2
P4	0	1	0	0	4
P5	0	1	1	1	16
P6	1	0	0	0	8
Validity of cues					
	0.80	0.875	0.75	0.625	

Note: The criterion was computed by the sum of the cue values multiplied by the weights 8, 4, 2, and 1 for the cues A, B, C, and D, respectively, with the exception of cue profile P5, to which the criterion value 16 was assigned. The criterion value plus a constant value of 3 was used as feedback in Experiment 2.

Table 2
The cue structure used in the test phase of the experiments and the predictions of how many times the models would select each patient as the most ill for all possible 21 pair comparisons

Patient	Cues (symptoms)				Criterion	Models' predictions				
	A	B	C	D		TTB_A	TTB_B	$COMP_A$	$COMP_B$	PROBEX
T1	0	0	1	1	3	0	0	0	0	3
T2	0	1	0	1	5	1	4	1	3	4
T3	0	1	1	0	6	2	5	2	5	6
T4	1	0	0	1	9	3	1	3	1	0
T5	1	0	1	0	10	4	2	4	2	1
T6	1	0	1	1	11	5	3	6	4	5
T7	1	1	0	0	12	6	6	5	6	2

Note: The criterion is computed by adding up the number of positive symptoms, multiplied by the weights 8, 4, 2, and 1 for the cues A, B, C, and D, respectively. Predictions A and B (represented with subscripts) refer to learned cue orders A, B, C, D and B, A, C, D, respectively, for both TTB and COMP.

2. Experiment 1

Experiment 1 tested the exemplar-based approach against the strategy-based approach for inferences from memory. The exemplar model PROBEX was tested against the strategies TTB, COMP, and TALLY. In contrast to the previous studies, such as Bröder and Schiffer (2003), we did not provide the relative importance of the cues to the participants. Instead, the participants had to learn them by feedback. This is a key difference because knowing cue importance is a precondition for applying inference strategies, thus the conclusions from the previous research might be limited by the fact that cue importance was provided to participants (for this critique see also Juslin & Persson (2002) and Dougherty et al. (2008)). As described above the experiment was conducted in three phases. To allow a proper test of the different models it was crucial to ensure that participants were able to retrieve all cue values; that is, in the first memorization phase of the experiment the cue values had to be nearly perfectly memorized. Otherwise, in the following training phase the participants would not be able to learn the relative importance of the cues when they received feedback about which of the two alternatives was the correct one. The knowledge acquired in the training phase could be used in the final test phase, where the models were tested with pairs of alternatives that were not used in the training phase.

Compared to the previous research, our experimental approach has several advantages. First, the inferences depended solely on the information retrieved from memory. Second, knowledge about the relationship between the cues and the criterion had to be learned, and was not provided to the participants, giving no a priori advantage to the strategy-based approach. Third, the crucial generalization test of the models was performed with pairs of alternatives that had to be retrieved from memory and for which no feedback had previously been provided.

2.1. Method

2.1.1. Participants

Twenty-five people with an average age of 25 years participated in the experiment. The participants were mainly students from various departments at the Free University of Berlin. Payment

[2] We performed a post hoc sensitivity analysis for the data of Experiment 2 to see whether the accuracy of the exemplar model depends on the simplifying method of setting the similarity parameter to a fixed value of $s = 0.5$. This analysis revealed that the best value for the similarity parameter fitted to each individual in both experiments was on average $s = 0.48$ (SD = .08). But even with this fitted free parameter the model was not substantially better in predicting participants' inferences (the fitted exemplar model could predict 0.4% more inferences). Thus, our conclusions do not depend on the simplifying restrictions for the exemplar model.

depended on the participants' performance. On average each participant received 24 euros and the average time spent on the experiment was 79 min.

2.1.2. Procedure

The cover story for the experiment was that a group of patients was suffering from a tropical fever with four distinct and easily identifiable symptoms. Similar tasks have been used before (e.g., see Garcia-Retamero, Hoffrage, and Dieckmann (2007), Gluck and Bower (1988), and Koehler (2000)). The fever had many different stages and the task of the participants was to judge which of two patients had reached the more advanced and dangerous stage. This task is not intended to represent real medical diagnoses, but nevertheless it appears more realistic than completely artificial tasks and should therefore facilitate the memorization of patients' symptoms. In the memorization phase, participants first had to memorize the presence or absence of the four symptoms for each of 13 patients. Each patient was given a German first name and had a unique set of symptoms. In the following training phase, participants compared the patients and predicted which of the two patients was in a more dangerous stage. After each prediction participants received feedback about whether the inference was correct. Finally, in the test phase of the experiment participants compared another set of patients whose symptoms they had memorized in the memorization phase but whom they did not encounter in the training phase. This phase represents the crucial test of whether the participants were able to generalize the acquired knowledge of how symptoms are related to the stages of the diseases to assess patients they had not yet encountered. Six of the 13 fictitious patients were used for training and the remaining 7 were used in the test phase (see Tables 1 and 2). Participants were paid according to their performance, so that good memorization and correct inferences in the test phase were rewarded. The participants were provided with two sets of written instructions: The first was read before the memorization phase and the second before the training phase (see Appendix).

The procedure for memorizing symptoms was similar to the method used by Bröder and Schiffer (2003). One difference was the use of monetary incentives for good performance. All 13 patients were repeatedly presented in blocks in which they appeared in a random order. Seven blocks of practice items alternated with seven blocks of items for which participants were paid, such that a practice block always preceded a paid block. Practice blocks were intended to give the participants an opportunity to learn the symptoms without monetary consequences, whereas paid blocks provided monetary incentives. Given the name of a patient, the participant had to predict (recall) for each symptom whether it was present or absent. In a practice block the name of a patient along with a choice of two symptoms, for example, "no headache–headache" was displayed and the participants selected one symptom by pressing the appropriate arrow button on the keyboard. The order of the two symptoms was randomized, so that the motor response (pressing the left or right key on the keyboard) could not be encoded as a proxy for remembering the correct symptoms. Thereafter the correct answer, for example, "headache", replaced the text line. If the wrong answer was given, asterisks were added on both sides of the correct symptom. This procedure continued until all symptoms (and potential asterisks) were visible. If all symptoms were correctly recalled the participant was allowed to proceed with the next patient by pressing the space bar. Otherwise the same procedure was repeated for the same patient again, which guaranteed that every patient would be correctly memorized at least once during each training block.

In the paid blocks, participants earned (lost) 4 eurocents for each symptom they recalled correctly (incorrectly). The procedure differed from the practice block in that after completing one patient a new patient appeared regardless of any errors in recall. In addition, only the patient's name and the current symptom choice, instead of all symptoms, were shown, to aid remembering each symptom. Feedback of either "right" or "wrong" was shown for the symptoms processed so far but not what the previous symptoms were to make each symptom choice in the paid blocks completely memory based. Thus, ideally, each symptom would later be recalled easily when cued by the name of the patient only. The total amount of money earned so far was shown on the computer screen at all times. All labels (patient names and symptom names) used in the experiment were randomized for each participant. Symptoms were always presented in the same order but the order was randomized for each participant. Participants were informed that if they did not finish all blocks within 100 min, they would be transferred to the next phase of the experiment (which, however, did not happen in the experiment). The time and number of paid blocks were displayed on the computer screen.

The second phase of the experiment—the training phase—used five blocks with 15 items each. An item consisted of a pair of patients, and the 15 items consisted of all possible pairs of patients from the set of six patients used in this phase (see Table 1). The procedure was simple. Two names were presented, and the participant had to select the patient who was most ill. The correct answer was then underlined and colored green when correctly chosen and red otherwise, along with a displayed text string of either "right" or "wrong." The instructions emphasized that this training phase should be used to develop a method that would allow the participants to make predictions in the test phase about which patient was in a more severe stage of the disease based on the memorized symptoms of the patients (see Appendix). The instructions did not include directions for what form this inference method for the final test phase should take. If participants aimed to apply a specific strategy they had to learn required cue validities in the training phase. The validities of the cues computed on the basis of the pair comparison of the training phase are presented in Table 1.

In the third phase of the experiment—the test phase—five blocks of 21 items based on all possible pair comparisons of the remaining seven patients were used (see Table 2). No feedback was given except that "Continue" appeared as soon as a response was recorded. Participants received 15 eurocents for a correct choice and they paid 15 eurocents for an incorrect choice. The payments were added to or subtracted from the total payment but a running total was not presented to the participants to avoid giving outcome feedback. What was considered a correct response was not obvious, because there was no "a priori correct" way to generalize from the training phase to the test phase. Therefore, the criterion value for each patient was determined by multiplying the cue values for the symptoms A, B, C, and D, with the weights 8, 4, 2, and 1, respectively, and taking the sum across all weighted cue values (see Table 2). The patient with the larger criterion value was considered as the correct response. However, since no feedback was provided to the participants in the test phase, how performance was determined could not alter the decision process and is therefore irrelevant for testing the models. It was only necessary for determining participants' final payoffs.

The item set for the training phase was constructed in such a way that TTB and COMP made the same predictions and allowed high accuracy. When using the rank order A, B, C, D for the four symptoms (see Table 1), TTB allowed 93% correct inferences for all possible pair comparisons. One cue profile, namely, P5, was designed to be an exception for which TTB would make an error. P5 had the cue values 0, 1, 1, 1 for the symptoms A, B, C, and D, respectively, and the highest criterion value. However, since the cue profile P6 with the cue values 1, 0, 0, 0 had a positive cue value for symptom A, TTB incorrectly predicts that this exemplar has a higher criterion value than P5. To avoid this mistake TTB could use an alternative rank order for the cues, namely, B, A, C, D, which would also correspond to the cues' validities. However, with this rank order TTB incorrectly infers that P4 has a higher criterion value than P6. Thus both cue orders lead to the same accuracy. For this reason we test two rank orders for TTB, namely, A, B, C, D, referred to as TTB_A, and B, A, C, D, referred to as TTB_B. Both versions of TTB were used to predict participants' inferences in the final test phase of the experiment.

For the compensatory strategy COMP we used the weights 6, 4, 3, and 2 for the symptoms A, B, C, and D, respectively. Similar to the procedure for TTB, two different weight orders were used for the symptoms A, B, C, D, namely, 6, 4, 3, 2 ($COMP_A$) and 4, 6, 3, 2 ($COMP_B$). In the training phase $COMP_A$ ($COMP_B$) makes the same predictions as TTB_A (TTB_B), so that in this phase both types of strategies perform equally well by making 93% correct inferences. For the test phase the weights 6, 4, 3, and 2 led to predictions for COMP that were very similar to the predictions of TTB. If we had used more similar weights (e.g., 4.5, 4, 3, and 2) then COMP's predictions would have been more distinct from TTB's predictions. In this situation someone using COMP with distinct weights could have been misclassified as using TTB. Thus, the use of the distinct weights for COMP has the advantage that everyone who used less distinct and more equal weights was still classified as using the compensatory strategy COMP.[3]

The item set for the training phase was constructed in such a way that the exemplar model made predictions in the test phase that differed from the prediction of the strategies TTB and COMP. The exemplar model performed perfectly in the training phase, because unique cue profiles with different criterion values were used. The item set of the test phase was constructed such that the exemplar model, TTB, and COMP made different predictions. The exemplar model predicted that the cue profiles T1, T2, and T3, which are similar to the cue profile P5 (0, 1, 1, 1) of the training phase, would be judged to have a high criterion due to the high criterion value of P5. In contrast, the strategies TTB and COMP would treat profiles similar to 0, 1, 1, 1 very differently, depending on the weights used by COMP and the cue order used by TTB. Table 2 shows the models' predictions and illustrates that the strategies make rather different predictions, which also differ from the predictions of the exemplar model. The predictions of the strategy TALLY were not considered for constructing the item sets and are not displayed in Table 2. For 50 of the 75 pair comparisons in the training phase and for 75 of the 105 pair comparisons in the test phase TALLY made an ambiguous prediction; that is, the two compared alternatives had the same sum of positive cue values. For those items of the training and test phase where it makes an unambiguous prediction, TALLY predicted the same choice as the COMP strategy.

2.2. Results

The results are presented in two parts. First a descriptive overview of the participants' performance is given. Second, we examine how well the various models predict participants' inferences.

2.2.1. Descriptive analysis

The memorization phase was very effective. As early as in the first paid block of the memorization phase the average proportion of correctly remembered symptoms was 73% (SD = 16%), which is clearly above chance performance. In the final paid block this average proportion increased to 97% (SD = 6%). Of 25 participants, 17 remembered all patients 100% correctly in at least one block. Participants got faster in recalling the patients' symptoms during the memorization phase. The first practice block took an average of 11 min compared with the last practice block, which took 2 min. The first paid block took an average of 5 min compared to the final block with 2 min. The practice blocks took longer than the paid blocks because any mistake in a patient's symptoms led to the repetition of the patient. In total the participants spent on average 48 min in the memorization phase.

In the training phase participants quickly learned to pick the correct patient (according to the corresponding feedback). As early as in the first block the participants performed on average above chance level with 81% correct answers (SD = 9%). In the final block the participants made an average of 96% (SD = 5%) correct inferences. In the training phase, one way of solving the task was simply to rank the patients in memory and to ignore all memorized cues, although this would lead to bad performance in the test phase. Using the rank of patients would have allowed a very quick decision. It is therefore illuminating that the median response time, that is, the time participants required to make an inference, in the last block was 3.3 s per item, indicating that participants did more than simply recalling patients' ranks. Instead, participants presumably also recalled additional information that would allow them to make an inference without knowing the ranks. Participants in general spent little time observing the given feedback. The median time between receiving the feedback for an inference and commencing with a new inference decreased from 1.8 s in the first block to 0.47 s in the last block.

In the final test phase participants made their inferences without receiving any feedback. The median response time for their inferences was 4.4 s per item in the last block. This is only slightly slower than in the training phase, which is consistent with the idea that the inference process is the same for the participants in both phases.

2.2.2. Predicting participants' inferences

Table 3 shows the average percentage of predicted inferences for the different models. In general, the strategies TTB and COMP predicted participants' inferences much better than the exemplar model. PROBEX, on average, predicted only 55% of all inferences, which does not differ substantially from what can be expected from pure chance ($p = .16$ according to a one-tailed binomial test). In contrast, the strategies TTB and COMP predicted the inferences substantially better than pure chance. In particular, the inference strategies TTB_A and $COMP_A$ predicted most participants' inferences,

Table 3
Average percentage of predicted inferences (*M* and SD) and average QS fit according to the quadratic scoring rule for all participants

Model	Experiment 1			Experiment 2		
	M (%)	SD (%)	QS fit	*M* (%)	SD (%)	QS fit
TTB_A	82	17	0.22	83	13	0.25
TTB_B	69	14	0.39	63	13	0.43
$COMP_A$	82	16	0.23	85	12	0.23
$COMP_B$	73	12	0.36	67	12	0.41
TALLY	60	4	0.42	61	4	0.39
PROBEX	55	10	0.47	49	9	0.48

Note: QS fit refers to the average score of the quadratic scoring rule, with lower values indicating better fit.

[3] We did an exhaustive search of weighting schemes for thecompensatory strategy under the constraint that all weights had to be different and follow a decreasing order for the cues A, B, C, and D. We did this by using integers between 1 and 30 as weights for the four cues. This provided us with 40,920 different weighting schemes. For all possible weighting schemes we determined the predictions of the compensatory strategy for the items of the test phase. This search showed us that there are only five possible patterns of predictions for a linear model and only one of them leads to decisions that are in line with the predictions of TTB (this weighting schema was observed 5068 times out of the remaining 36,456 weighting schemes after removing those that led to ambiguous predictions).

which indicates that most participants gave larger relative importance to symptom A (note that the labels for the cues were randomized across participants, so that for the majority of participants cue A was not given the label "A"; therefore the labels do not explain why cue A was given more importance). TALLY does not predict a large percentage of inferences, but this is mainly due to the large number of items for which TALLY made ambiguous predictions. For those items where TALLY predicted a random choice we counted .50% points.

As the crucial goodness-of-fit measure we evaluated the strategies' predictions with the quadratic scoring rule (see, for example, Selten (1998)). For this purpose we applied a naïve error theory for every strategy, which assumes that a strategy is not perfectly applied but that, with a probability of ε, an error occurs and the unpredicted alternative is chosen. Thus, a strategy predicts the choice of an alternative with a probability of $1-\varepsilon$ (for different error concepts see also Mata, Schooler, and Rieskamp (2007), Rieskamp (in press), and Rieskamp, Busemeyer, and Mellers (2006)). For each item we determined the sum of squared differences of the predicted probability of choosing each of the two alternatives and the observed behavior. Accordingly, a strategy i with application error ε reaches a fit QS of

$$QS(i,\varepsilon) = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{2}[D_t(k) - p(k|i,\varepsilon)]^2 \tag{3}$$

with t for the item, T as the number of items, and k for the two alternatives; $p_t(k|i,\varepsilon)$ denotes the probability of choosing alternative k when using strategy i, and $D_t(k)$ is an indicator function that equals 1 if alternative k was chosen and equals 0 if the alternative was not chosen. For each strategy and participant the optimum value for the application error ε was selected that led to the lowest, that is, the best value for QS. The quadratic scoring rule provides a mean sum of squared errors that ranges from 0 to 2, with 0 as an optimal fit where the observed choices are predicted with probability 1. The quadratic scoring rule has the advantage that it will give a better goodness-of-fit to TALLY than to COMP for someone who is using TALLY.[4]

Table 3 shows that on average TTB predicted the inferences best with an average fit of QS = .22, which was similar to COMP with an average fit of QS = .23. TALLY, which predicted a random choice for many pair comparisons did not on average reach a good fit. In sum, although the quadratic scoring rule can lead to different results from those obtained with the percentage of predicted choice, in our case the conclusions that can be drawn from the two measurements are the same.

To gain more insight on which model predicted the inferences best, we assigned each participant to the model with the best fit (i.e., the lowest QS score). Because we simply made this assignment (similar to what Bröder and Schiffer (2003) did) without taking the magnitude of the difference of the strategies' fits into account, this does not provide us with an estimate of how probable it was that the best-fitting strategy was actually selected by a particular participant. Therefore, we will only interpret the strategy assignment for the whole sample of participants and test it against an assignment that would result from pure chance.

Table 4
Number of participants assigned to the different models, their respective percentage of predicted inferences (*M* and SD), and their QS fit according to the quadratic scoring rule

Model	Experiment 1				Experiment 2			
	Assigned	*M* (%)	SD (%)	QS fit	Assigned	*M* (%)	SD (%)	QS fit
TTB	15	95	4	0.12	6	94	4	0.11
COMP	6	92	6	0.14	16	92	6	0.20
TALLY	3	61	2	0.40	3	63	2	0.38
PROBEX	1	84		0.27	0			

Note: QS fit refers to the average score of the quadratic scoring rule, with lower values indicating better fit.

Table 4 shows the number of participants assigned to each model. The majority of participants (15 out of 25) were assigned to TTB (with 12 to TTB_A and 3 to TTB_B), which is much larger than expected from pure chance assuming equal frequencies for all four strategies, $\chi^2(3) = 18.36$, $p < .01$. Most of the remaining participants were assigned to COMP or TALLY and only one participant was assigned to PROBEX. These results provide clear evidence that people rely on strategies to solve the inference problem and do not follow an exemplar-based inference process. Moreover, most participants, in line with the results of Bröder and Schiffer (2003), seemed to rely on the simple lexicographic heuristic TTB when making inferences from memory.

2.3. Discussion

The first experiment aimed to test the strategy-based approach against the exemplar-based approach in predicting people's inferences. This test led to a clear result: The exemplar model was not able to predict participants' inferences from memory. In contrast, the strategy-based approach was successful in predicting a substantial proportion of inferences. The simplest strategy, the lexicographic heuristic TTB, was best in predicting participants' inferences. When we classified participants' inference strategies, those participants who were not assigned to TTB used a compensatory strategy. These results are in line with the results of Bröder and Schiffer (2003). This is an important result, because in our experiment participants were not provided with any information about the validity or relative importance of the cues, in contrast to Bröder and Schiffer's study. Gigerenzer and Goldstein (1996) predicted that it is more likely to find people using TTB when information has to be *searched from memory* compared to when *searched from givens*. Our findings are consistent with this prediction. TTB predicts people's inferences from memory well. It is interesting that we find this in a situation in which the cue validities were not given to the participants (as in many previous studies), but had to be learned from feedback. Recently, Dougherty et al. (2008) criticized that providing participants with cue validities does not provide a strong test of the TTB heuristic. The present study therefore takes up this criticism and shows that TTB predicts participants' inferences well in a situation in which all information has to be memorized and a subjective cue hierarchy has to be learned (see also Gigerenzer et al. (2008)). The support for TTB and also for the COMP strategy suggests that many participants learned to distinguish the cues according to their predictive accuracy and identified the most valid ones.

One reason why the exemplar model was not very successful in predicting inferences could have been the type of feedback in the training phase, in which the only information provided was

[4] To demonstrate the quadratic scoring rule, consider a person who applies the strategy TALLY with an application error of 20%. In the test phase with 105 items, TALLY predicts on average 80% of the choices for those 30 items where it makes an unambiguous prediction and for the remaining 70 items TALLY predicts a random choice, so that when counting .50 percentage points TALLY will on average predict 59% of the choices. The strategy COMP will on average predict the same percentage of choices, because COMP predicts the same choices for those items where TALLY makes an unambiguous prediction and for the remaining items COMP will on average predict 50% of the choices. Thus, by using the percentage of predicted choices it is not possible to identify that the person applied TALLY. In contrast, when using the quadratic scoring rule TALLY has an expected fit of QS = .45 whereas COMP has an expected fit of only QS = .55. This is due to the low fit of COMP for those items where the person using TALLY makes random choices. In contrast, for a person using COMP with an application error of 20%, the quadratic scoring rule will accurately give COMP an expected fit of QS = .32, which is better than TALLY's expected fit of QS = .45.

whether the choice was correct or incorrect. Participants did not receive information about the severity of the disease; that is, the criterion value was not provided. However, to apply the exemplar model correctly it was necessary to assign at least an ordinal rank describing the severity of the disease. Because all possible patient pair comparisons of the training phase were shown these ranks could easily be inferred. That is, after some time participants could easily detect that a particular patient was in the most severe stage of the disease. Nevertheless, because these necessary ranks had to be inferred by the participants if they were to follow an exemplar-based inference process, this process might have been too demanding in comparison to the use of a simple inference strategy, such as TTB. Because inferring a criterion value for each patient in the training phase was necessary for the exemplar model but not necessary for the strategy TTB, this might have fostered a strategy-based inference process. Therefore, providing explicit numerical feedback about the criterion for single objects could make an exemplar-based inference process more likely, which we tested in Experiment 2.

3. Experiment 2

The nature of feedback ought to be important for how people represent information. In the second experiment feedback was provided that not only indicated the correct choice, but additionally provided information about the criterion value of each alternative. If memorized patients are associated with an explicit criterion, the feedback might facilitate an exemplar-based inference process. This type of feedback might also foster inference processes in line with the compensatory strategies. The compensatory strategies compute a score for each alternative and select the alternative with the highest score. The additional feedback about the alternatives' criterion values could be used to evaluate the scores the compensatory strategies produce (Juslin, Jones, et al., 2003). If the scores do not match the feedback, then the weights used for the compensatory strategies could be adjusted to improve the accuracy of the strategies. In contrast, the additional continuous feedback does not improve the accuracy of TTB, because TTB does not provide any estimates about the alternatives' criterion values. Thus, the additional feedback about the alternatives' criterion values could make an exemplar-based inference process more likely, or alternatively, lead to a more frequent application of compensatory strategies. This prediction was also tested in Experiment 2.

3.1. Method

3.1.1. Participants

Twenty-five people with an average age of 26 years participated in the experiment. The participants were mainly students from various departments at the Free University of Berlin. Payment depended on the participants' performance. On average each participant received 21 euros and the average time spent on the experiment was 84 min.

3.1.2. Procedure

The procedure of Experiment 2 was very similar to that used in Experiment 1, with the only difference being that feedback about the alternatives' criterion values was additionally provided in the training phase. As feedback, a constant of 3 was added to the criterion values presented in Table 1. Experiment 2 employed similar instructions to those used in Experiment 1 (see Appendix). The only difference was that the following sentence was added to the instructions explaining the training phase: "in addition, you will be provided with the percentage of viruses in the blood for both patients relative to the percentage that is lethal. A patient with a higher percentage is in a later stage of the disease."

3.2. Results

The results are presented in three parts. First, a descriptive overview of participants' performance is given. Second, we examine how well the various models predict participants' inferences. Finally, we examine participants' response times and relate them to the models.

3.2.1. Descriptive analysis

The average proportion of correctly remembered symptoms in the first block of the memorization phase with paid items was 67% (SD = 14%), which increased to 96% (SD = 7%) for the final block. Of 25 participants, 15 remembered all patients 100% correctly in at least one block. The first paid block was completed in 4 min on average, and the final block was completed in 2 min. The total time spent in the memorization phase was 47 min on average. The first block of practice took on average 10 min to finish, and the final block was finished in 2 min.

In the training phase the participants quickly learned to infer which of two patients was in a more severe stage of the disease. Participants were already above chance level in the first block with a proportion of 83% correct answers (SD = 11%), which increased to 97% (SD = 5%) in the final block. The median response time in the last block was 4.0 s, which is slightly longer than in Experiment 1. The median time spent observing the feedback was 3.1 s in the first block and only 0.7 s in the last block. This is also more than in Experiment 1 (most notably for the first block with 1.8 s). In the test phase the participants had a median response time of 4.1 s for the last block, which is the same as in Experiment 1.

3.2.2. Predicting participants' inferences

Table 3 shows the average percentage of predicted inferences of the different models. In general, all inference strategies were able to predict participants' inferences much better than the exemplar model. Again, as in Experiment 1, the fit of PROBEX with 49% does not differ from pure chance. In contrast, all other models were able to predict the inferences substantially better than chance level. In particular, as in Experiment 1, the inference strategies TTB_A and $COMP_A$ predicted the most participants' inferences, which indicates that most participants gave larger relative importance to symptom A. However, in contrast to Experiment 1, the strategy $COMP_A$ predicted more inferences than TTB_A.

As the crucial goodness-of-fit measure we again evaluated the strategies using the quadratic scoring rule. Table 3 shows that on average, $COMP_A$ predicted the inferences best with an average fit of QS = .23, which was similar to TTB_A with an average fit of QS = .25. Interestingly, TALLY achieves a better fit in comparison to the remaining strategies according to the QS measure; that is, TALLY has a better QS than $COMP_B$ and TTB_B, which differs slightly from the results on the percentage of predicted inferences reported above. Nevertheless, the quadratic scoring rules suggests that $COMP_A$ and TTB_A do best in predicting participants' inferences.

As in Experiment 1, the inference model that had the best fit according to the quadratic scoring rule was assigned to each participant. Table 4 shows the number of participants assigned to each model. In contrast to Experiment 1, the majority of participants (16 out of 25) were assigned to $COMP_A$ (none to $COMP_B$), which is much larger than expected from pure chance, $\chi^2(3) = 23.16$, $p < .01$. Six participants were assigned to the compensatory strategy TALLY. The remaining participants were assigned to TTB (with 4 to TTB_A and 2 to TTB_B). None of the participants were assigned to PROBEX.

These results provide clear evidence that people rely on strategies to solve inference problems and do not follow an exemplar-based inference process. However, in contrast to Experiment 1 and to the results of Bröder and Schiffer (2003), when our participants received feedback about the alternatives' criterion values, the majority seemed to rely more frequently on compensatory strategies. Thus, when making inferences from memory people do not select simple strategies per se, such as TTB, as previous research has suggested; instead, people could still be applying compensatory strategies.

3.2.3. Examining participants' response times

Although our main conclusions are based on the analysis of participants' inferences, response times are an important criterion on which to evaluate cognitive models of decision processes (Luce, 1986). For instance, for those participants applying TTB one would expect faster inferences the earlier a cue discriminates (Bröder & Gaissmaier, 2007; Persson, 2003). For participants applying COMP and TALLY, the response time should be a function of the difference between the score determined for one alternative and the score determined for the other (see also Bergert and Nosofsky (2007)). The underlying rationale for this prediction relies on the idea that people will not literally execute a strategy step-by-step when its prediction can easily be foreseen. For example, if one alternative has only positive cue values and the other has only negative cue values, someone applying COMP does not need to determine the exact score for each alternative. Instead the person can foresee COMP's prediction and makes a corresponding choice. In contrast, if some cues favor one alternative and some cues favor the other, the decision maker cannot simply foresee COMP's prediction and has to compute the alternatives' scores, which will take some time. Accordingly, the larger the difference between the alternatives' scores the faster a decision should be made. In contrast, when TALLY is applied and both alternatives have the same score, a slow response time is expected. For the simplified version of PROBEX it can be predicted that the response times should not differ across items, because the same number of exemplars always needs to be retrieved from memory to make an inference. Likewise, Juslin and Persson (2002), who used the full version of PROBEX and estimated the model's parameter on the basis of experiment data, also found no response time difference for various cue profiles.

Because neither Experiment 1 nor Experiment 2 was specifically designed to test the models' response time predictions, we examined the response times for groups of items. Additionally, we merged the data from the two experiments to obtain larger sample sizes. When examining the response times for those participants assigned to TTB, we considered all items where the first most valid cue discriminated compared with all other items where the first cue did not discriminate. Participants assigned to TTB should be faster for the first group of items than for the second group. For each participant we first determined the median response time for each group of items and then determined the average median response time across all participants for the two groups. As predicted, for the 21 participants assigned to TTB, the average median response time for the first group of items was 8.1 s (SD = 6.05) compared with 9.6 s (SD = 3.67) for the second group ($Z = 2.5$, $p = .014$, according to a Wilcoxon test).

When examining the response times for participants assigned to COMP or TALLY, we considered all items where the sum of cue values differed for the two alternatives so that TALLY could discriminate between the alternatives in comparison to those items where the cue sums were the same and TALLY could not discriminate. For the first group of items the average difference between the two scores determined by COMP for the alternatives was 3.5 as opposed to the second group with an average difference of 2.3. Therefore, participants assigned to TALLY or COMP should have a faster response time for the first as opposed to the second group of items.

In line with these predictions, the 6 participants assigned to TALLY made their inferences with an average median response time of 3.6 s (SD = 2.0) for the first group of items as opposed to an average median response time of 5.7 s (SD = 2.8) for the second group ($Z = 2.2$, $p = .027$). Note that the response times for participants using TALLY were much faster than for those assigned to TTB. This is surprising, since given the definition of TALLY and TTB with serial search processes (Gigerenzer & Goldstein, 1996), TTB searches on average for fewer cues and therefore one would expect people to apply TTB faster than TALLY. To explain this surprising finding one could argue that TALLY could be processed rather quickly if one does not assume a serial search process, as we will discuss in detail in the General Discussion below. Likewise, the 22 participants assigned to COMP made their inferences with an average median response time of 5.7 s (SD = 3.7) for the first group of items as opposed to an average median response time of 8.4 s (SD = 3.0) for the second group ($Z = 3.0$, $p = .002$). The response times for participants assigned to COMP were larger when compared to participants assigned to TALLY, which can be expected when considering that COMP requires a specific weighting of cues that TALLY ignores. Thus, overall the response times of the participants assigned to the different strategies were consistent with the strategies' predicted response times.

3.3. Discussion

The purpose of Experiment 2 was to examine whether people follow an exemplar-based inference process more frequently when continuous feedback about the criterion values of alternatives is provided, which is information required by the exemplar model. The results were clear-cut: Providing continuous feedback did not foster exemplar-based inference processes. PROBEX, the exemplar model, was not able to predict the inferences better than chance level. Therefore, Experiment 1's results—that people seem to follow a strategy-based inference process for inferences from memory—can be generalized to a situation with a continuous feedback. However, providing continuous feedback did alter the observed inference process: The inference process of the majority of participants was most often best predicted by the compensatory strategies. These results are in line with the work of Juslin, Jones, et al. (2003) and Juslin, Olsson, et al. (2003), who showed that for a task in which participants had to make single-object judgments, they relied on compensatory strategies more frequently when continuous feedback for the individual objects opposed to dichotomous feedback on the level of pair comparisons was provided. Finally, our analyses show that the response times for those participants in Experiments 1 and 2 assigned to a particular strategy were consistent with the expected response times for the strategy. Similarly, Bergert and Nosofsky (2007), as well as Bröder and Gaissmaier (2007) have also successfully used response times to analyze probabilistic inferences (see also Persson (2003)).

4. General discussion

In everyday decision making we rarely have all the relevant information laid out in a convenient matrix in front of us, as in experiments using the information board paradigm (Huber, Wider, & Huber, 1997; Payne et al., 1993; Rieskamp & Otto, 2006). The motivation behind using an information board is that it gives the researcher a high degree of experimental control over the information that is available for making a decision, making it easier to deduce the process that leads to a decision. However, the experimental results obtained using information boards limit generalizations to inference situations in which information has to be retrieved from memory. For this reason the study of inferences from memory is necessary, although such inferences present a serious methodological challenge to control for the memory content. Only when assumptions about what information has been stored in memory are justified can conclusions about different cognitive processes underlying inferences from memory be made. The primary goal of our study was to compare different models for predicting inferences from memory. Two main theoretical approaches were contrasted, namely, the strategy-based approach and the exemplar-based approach.

4.1. The failure of the exemplar-based approach

The idea of an exemplar-based inference process did not find any support in the two experiments that we conducted. This result was very surprising for us, considering how successfully exemplar models have been applied in the past to describe memory-based inference processes (Juslin, Jones, et al., 2003; Juslin, Olsson, et al., 2003; Juslin and Persson; 2002; Nilsson et al., 2005). The simplest explanation is that the exemplar-based inference process does not take place in a pair-comparison task because the nature of the task invites cue-wise processing. In the same vein, Rieskamp and Otto (2006) argued that support for the exemplar models in pair-comparison tasks used by Juslin, Jones, et al. has been provided in experiments in which the training was done with a single-object task, which makes it unlikely that strategies such as TTB will later be adopted in the final pair-comparison test.

There are other reasons for not following an exemplar-based inference process. First, exemplar models have hitherto mainly been supported in tasks using visual stimuli; the cognitive mechanism may thus not be compatible with objects that are presented verbally. Bröder and Schiffer (2003) manipulated the representational format and found that TTB predicted participants' inferences to a smaller extent with pictorial stimuli, but they did not explicitly test exemplar models as an alternative. Persson (2003) found that changing text-based stimuli into strings of four letters, for example "GHGG", made encoding of exemplars and use of them much easier, despite the abstract presentation format.

Second, the amount of training may be crucial for learning an exemplar strategy. Johansen and Palmeri (2002) found a representational shift from simple strategies to exemplar processes in categorization when participants had extensive training in the task. Consistently, Nosofsky and Bergert (2007) found that when a lot of training was provided people's inferences could be best described by an exemplar model. However, the authors suggest that an inference strategy that takes the configuration of cues into account could also explain the results. Evidence for the use of cue configuration strategies was also found by Garcia-Retamero, Hoffrage, Dieckmann, and Ramos (2007). Increasing the amount of training in our experiments would probably not lead to an exemplar-based inference process, because there were so few pairs. If the same pair comparison was presented repeatedly, participants would probably at some point simply retrieve the answer encoded on earlier trials associated with the shown names and thus effectively skip the retrieval of cues or exemplars from memory.

Third, working memory might be a bottleneck for following an exemplar-based inference process. The most common interpretation of an exemplar model is that it describes automatic cognitive processing of information (Johansen & Palmeri, 2002). Yet it is also conceivable that decision makers explicitly reason about the previous exemplars and behave accordingly, but this deliberate reasoning process might be too demanding when working memory has to handle the test exemplars and the retrieved training exemplars simultaneously.

Fourth, the cues, presented as symptoms, were easy to memorize, but this might also lead to a specific representation that later either hindered or fostered the selection of specific strategies. For example, participants could have memorized patients by encoding sentences such as "this is Peter. Peter has two symptoms. The symptoms are a rash and a headache". If this representation is retrieved verbatim it is obvious that TALLY is a strategy that could easily be applied, because the number of cues could be retrieved before the cues themselves are retrieved. This would also explain why people applying TALLY gave faster responses than people who applied TTB. In sum, although our results suggest that people do not follow an exemplar-based inference process for pair-comparison tasks, under specific situations where the cues are non-verbal and many training exemplars are provided, an exemplar-based inference process might occur.

4.2. The strategy-based approach

In line with the previous research (Bröder & Schiffer, 2003; Rieskamp & Otto, 2006), we found that the inference strategies TTB and COMP were best in predicting individuals' inferences. Bröder and Schiffer argued that the selection of TTB should occur more frequently when inferences are made from memory, because the information retrieval is cognitively demanding and fosters the selection of strategies such as TTB that require little information (see also Gigerenzer and Goldstein (1996)). In line with this assumption, in Experiment 1 TTB was the best model for predicting participants' inferences and most participants were classified as selecting TTB. However, interestingly and in contrast with this assumption, in Experiment 2 where explicit numerical feedback was provided, the strategy COMP predicted participants' inferences best. These results indicate that strategy selection is very sensitive to the nature of feedback. In sum, a memory-based inference process can lead to the selection of a simple heuristic such as TTB, but depending on the type of feedback people could also apply compensatory strategies. These results are novel, because the results of Bröder and Schiffer would lead one to conclude that simple inference strategies are predominately selected when inferences have to be made from memory. However, this interpretation has to be taken cautiously, because the strategies TTB_A and $COMP_A$ made different predictions for only one critical pair comparison. Since participants do not apply strategies without any mistakes only one critical pair comparison leaves room for misclassifications. However, this critical pair comparison was repeated five times, so that only in the unlikely event that a participant made an application error for the majority of these five pairs would a misclassification occur. Nevertheless, on a group level it was possible to analyze how often a strategy was best in predicting participants' inferences and whether this proportion for all participants differs from pure chance level.

The result that people select different strategies depending on the type of feedback is in line with the studies by Juslin, Jones, et al. (2003) and Juslin, Olsson, et al. (2003), who found that participants switched from an exemplar-based inference process to a strategy-based inference process when the type of feedback was changed from categorical to numerical. They argued that the way in which cues are cognitively represented strongly influences a judgment process. However, strategy selection could also be explained as a direct effect of how the feedback is provided. Rieskamp and Otto (2006) proposed the strategy selection learning (SSL) theory to explain when and why people select a particular strategy. In a nutshell, the SSL theory assumes that people have a repertoire of strategies and that they select the strategy they expect to be the most successful in solving the inference problem they face. When receiving feedback, people revise their evaluations of strategies so that the strategy that performs best in a particular environment becomes the strategy most likely to be selected.

This theory gained support in several experiments in which participants acquired information by using an information board. Although Rieskamp and Otto (2006) did not specify in detail how different types of feedback lead to different strategy reinforcements, the feedback used in Experiment 2 could lead to a more differentiated reinforcement process that favors compensatory strategies. In Experiment 1, many different weights for the compensatory strategy could have led to a good performance. In contrast, in Experiment 2 the weighted sum determined by the compensatory strategy could be compared with the numerical feedback about the alternatives' criterion values. When this feedback process leads to more reliable weights for the compensatory strategy, people might be more confident in using a compensatory strategy for new pair comparisons in the test phase. The additional feedback about the criterion values of the alternatives in Experiment 2 may also have had the effect that the participants were less likely to make explicit pair comparisons, which also could explain why few participants were assigned to TTB. Yet this cannot explain why no one was assigned to PROBEX, because if participants perceived the feedback on a single-object basis this should have increased the likelihood of an exemplar-based approach that requires this information.

4.3. Alternative approaches

The aim of our research was to test the strategy-based approach against the exemplar-based approach for predicting inferences from memory. Naturally, however, there are other competing approaches that are worth examining. Exemplar models represent an "all-purpose inference machine" (Juslin, Jones, et al., 2003, p. 925) because they can be used for various inference situations. Connectionist models, such as the adaptive network model proposed by Gluck and Bower (1988), are another type of all-purpose inference machine. Recently, Rieskamp (2006a) tested the strategy-based approach against the adaptive network model. Although both theories were able to describe participants' inferences, overall the strategy-based approach was more suitable for predicting people's behavior, especially in a dynamic environment.

Another all-purpose inference machine is represented by the sequential sampling approach (see, for example, Busemeyer and Townsend (1993), Koehler, White, and Grondin (2003), Lee and Cummins (2004), Newell (2005), and Wallsten and Barton (1982)). Sequential sampling models assume that people continuously sample pieces of evidence that accumulate over time. When the evidence for one option in comparison to the evidence for another option exceeds a threshold, the sampling process stops and the alternative with the most positive accumulated evidence is selected. When searching for cues in the order of their validity and when using a very low threshold, a sequential sampling model predicts choices consistent with TTB, and when using a high threshold the model predicts choices consistent with COMP (cf., Lee & Cummins, 2004).

Yet some of our results cannot easily be explained by a sequential sampling model. First, adjusting the threshold allows the sequential sampling model to make predictions in line with TTB and COMP, but to make predictions in line with TALLY, the weights for the cues also need to be changed. Second, referring to the response time data, when predicting choices in line with TALLY a higher decision threshold is needed than when predicting choices in line with TTB. A higher decision threshold should lead to slower decisions compared to a lower threshold. Therefore, the model predicts a slow response time for those participants assigned to TALLY (corresponding to a high threshold) and a fast response time for those participants assigned to TTB (corresponding to a low threshold). Yet the experimental evidence shows the opposite: Participants whose inferences could be best predicted by TALLY were much faster than participants whose inferences were best predicted by TTB. Nevertheless, since our study was not designed to test the strategy-based approach against the sequential sampling approach, this counterevidence should be interpreted carefully.

How can we explain the fast response time for TALLY in comparison to TTB? One explanation refers to the way information is represented in memory. In our experiments the participants had to memorize the symptoms of hypothetical patients. If these symptoms were memorized in a conjoint format, such as "Peter has two symptoms. The symptoms are a rash and a headache", the application of TALLY could be very fast, because the number of positive symptoms would be retrieved immediately. This representation of information for inferences from memory appears to be a plausible assumption. In contrast, for inferences from givens when using, for instance, an information board to present information, a slow serial information search process is necessary for TALLY. The important conclusion that can be drawn from this result is that the process predictions that can be derived from specific inference strategies depend often crucially on the way information is represented.

4.4. Adaptive decision making

The evidence from the two experiments shows that the strategy-based approach was most successful in predicting inferences from memory; the exemplar-based approach did not represent the inferences adequately. Surprisingly, though, there was no predominant simple strategy people selected when making inferences from memory. Instead, when people made inferences from memory different inference strategies were selected depending on the type of feedback.

The strategy people apply in any given situation depends on various factors, including time pressure, information search costs, the nature of feedback, and presentation format of information. Our study illustrates that humans can successfully cope with a challenging memorization task and effectively use their learned knowledge to make new inferences. The study supports the notion of an adaptive human decision maker.

Acknowledgements

We gratefully acknowledge helpful comments on previous versions of this article by Thorsten Pachur, Wolfgang Gaissmaier, and Arndt Bröder. We would like to thank Anita Todd and Julia Schooler for editing the manuscript.

Appendix Experimental instructions

The original instructions were in German and were formatted to enhance readability (e.g., additional line breaks were included). The instructions for Experiment 1 were given in two parts:

Part. 1

Welcome to the experiment. Please read the instructions carefully. The following study will last for about 2 h.

Overview of experiment: The experiment has two parts. In the first part you have to memorize the symptoms of 13 patients. You will get further instructions for the second part after you have successfully learned the symptoms.

First part, learning phase: In the first part you have to memorize the symptoms of 13 patients. The patients can show up to four possible symptoms: rash, headache, cough, and dizziness. One patient can have, for example, a rash and a headache but not a cough or dizziness. The descriptions of the patients will be presented repeatedly on the computer screen, to allow you to memorize the patients and their symptoms. You will only see one description for one patient at a time, and for each of the four symptoms you have to predict whether the patient shows the symptom or not. If you have predicted all four symptoms correctly, you will make the same predictions for the next patient. If you happen to make an error, you will be asked to make the predictions for the same patient once again. During the learning phase there will be seven test phases where you will be paid for every correct prediction. Your total earnings will be shown on the screen during the test phase. You will receive 0.04 euros for each correct answer, and 0.04 euros will be deducted for every wrong answer. You will thus not earn any money if you make random predictions. You can earn 16 cents for each patient out of the 13 patients if you predict the correct symptoms, which means that you can earn 2.08 euros in each of the seven test phases. We know that it is very difficult to memorize the symptoms of all patients. Therefore, do not be disappointed if it takes some time before you can correctly predict the symptoms. Still, you should try to memorize the symptoms as thoroughly as possible, since it will help you to earn more money later in the experiment.

There is a time limit of 1 h 40 min for the learning phase, to prevent the experiment from lasting too long. If you are not able to finish in time, you will automatically be transferred to the next part of the experiment. If you run out of time and have not learned the symptoms of the patients correctly, you will surely have a disadvantage in the next part of the experiment and will probably earn less money. You will probably need a lot of time initially. You will be faster after you have learned some symptoms. The time remaining to you will be shown after you have predicted each patient's symptoms. For every symptom, you should predict if the current patient has this symptom. Press the "left arrow" key if you want to choose the symptom description positioned to the left side of the screen and the "right arrow" key correspondingly for the symptom description positioned to the right. To proceed to the next prediction, you have to press the space bar after each patient. If you want to take a break to concentrate, it is best to do so before you press the space bar.

Thank you very much for participating in this experiment. If you still have questions, then read through the instructions again or ask the experimenter. All recorded data will of course be treated anonymously. Please do not take any notes and do not talk to other participants in the experiment.

Part. 2

Congratulations! You have successfully learned the symptoms of 13 patients. In the second part of the experiment you will compare the patients with respect to the stage of their disease. All 13 patients suffer from a tropical fever. The fever has reached a different stage of the disease for the 13 patients. Your task is to choose which of two patients is in a later and therefore more dangerous stage of the disease. You will make 180 decisions in total. This second part is divided into two phases (A and B). In each phase you will only compare half of the patients for whom you learned the symptoms. That is, in phase A you will not compare the same patients as in phase B.

Phase A: For the first 75 decisions you will only compare one half of the patients. The symptoms of the patients will not be shown to you. Thus, you have to rely on your memory. Since you have not learned anything about the stages of the tropical fever, in the beginning you will only be able to guess for which of the patients the disease is in a later stage. However, you will receive feedback about whether you chose the correct patient, that is, the patient in a later stage of the disease. You can improve your decisions with the feedback. It is helpful not to assume that the appearance of many symptoms necessarily goes with a later stage of the disease. For your first 75 decisions you will not get any payment, that is, you will not receive any money nor will you lose any money. For this reason it is not so important to make correct decisions. Instead it will be necessary to develop a method that you can use to compare patients with each other in respect to the stage of disease. In the following phase B you will compare the patients whom you did not see in phase A. Concentrate therefore less on the patients and more on the symptoms of the patients.

Phase B: For the last 105 decisions of the second phase you will compare the second half of the patients with each other in the second phase. In this phase you will not get any feedback about whether your decisions are right or wrong. That is, to be successful you have to use what you have learned up to this point. This includes the memorized symptoms for the patients and in addition what you have learned about these symptoms in phase A. It is therefore important for you to think about a method in phase A that can be used to make the remaining decisions in phase B. Note that in phase B you will compare patients with each other who were not compared with each other in phase A. You will receive 15 cents for each correct decision, and you will lose 15 cents for each wrong decision. Thus, you will not earn any money if you make random predictions. On the other hand if you make good predictions you will be able to earn a considerable profit. You will only be paid for your decisions in phase B. But it is only in phase A that you can get the knowledge necessary to make good decisions in phase B. Your earned payoff will not be shown during this part of the experiment, but it will be shown to you after the experiment.

References

Aha, D. W. (1997). Lazy learning. *Artificial Intelligence Review, 11*, 7–10.

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 107–129.

Bröder, A. (2000). Assessing the empirical validity of the "take-the-best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1332–1346.

Bröder, A. (2003). Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 611–625.

Bröder, A., & Gaissmaier, W. (2007). Sequential processing of cues in memory-based multi-attribute decisions. *Psychonomic Bulletin and Review, 45*, 895–900.

Bröder, A., & Schiffer, S. (2003). Take the best versus simultaneous feature matching: Probabilistic inferences from memory and effect of representation format. *Journal of Experimental Psychology: General, 132*, 277–293.

Busemeyer, J. R., & Townsend, T. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*, 432–459.

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

Dhami, M., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin, 130*, 959–988.

Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory and Cognition, 35*, 1801–1813.

Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review, 115*, 199–211.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32*, 53–88.

Garcia-Retamero, R., & Rieskamp, J. (in press). Adaptive mechanisms for treating missing information: A simulation study. *The Psychological Record*.

Garcia-Retamero, R., Hoffrage, U., & Dieckmann, A. (2007). When one cue is not enough: Combining fast and frugal heuristics with compound cue processing. *Quarterly Journal of Experimental Psychology, 60*, 1197–1215.

Garcia-Retamero, R., Hoffrage, U., Dieckmann, A., & Ramos, M. (2007). Compound cue processing within the fast and frugal heuristics approach in nonlinearly separable environments. *Learning and Motivation, 38*, 16–34.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650–669.

Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Fast and frugal models are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas. *Psychological Review, 115*, 230–239.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology, 52*, 464–474.

Gluck, M. A., & Bower, H. B. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227–247.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Huber, O., Wider, R., & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica, 95*, 15–29.

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

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

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

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.

Koehler, D. J. (2000). Probability judgment in three-category classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 28–52.

Koehler, D. J., White, C. M., & Grondin, R. (2003). An evidential support accumulation model of subjective probability. *Cognitive Psychology, 46*, 152–197.

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

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review, 107*, 227–260.

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin and Review, 11*, 343–352.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision, 52*, 29–71.

Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging, 22*, 796–810.

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

Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences, 9*, 11–15.

Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 53–65.

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical test of a fast-and-frugal heuristic: Not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes, 91*, 92–96.

Nilsson, H., Olsson, H., & Juslin, P. (2005). The cognitive substrate of subjective probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 600–620.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.). *Essays in honor of William K. Estes* (Vol. 1, pp. 149–167). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M., & Bergert, F. B. (2007). Limitations of exemplar models of multi-attribute probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 999–1019.

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous spaces. *Psychonomic Bulletin and Review, 5*, 345–369.

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

Persson, M. (2003). Bounded rationality and exemplar models. *Dissertation Abstracts International, 65*, 272 [Doctoral dissertation, Uppsala University, 2003].

Rakow, T., Hinvest, N., Jackson, E., & Palmer, M. (2004). Simple heuristics from the adaptive toolbox: Can we perform the requisite learning? *Thinking and Reasoning, 10*, 1–29.

Rapoport, A., & Wallsten, T. S. (1972). Individual decision behavior. *Annual Review of Psychology, 23*, 131–176.

Rieskamp, J. (2006a). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1355–1370.

Rieskamp, J. (2006b). Positive and negative recency effects in retirement savings decisions. *Journal of Experimental Psychology: Applied, 12*, 233–250.

Rieskamp, J. (2008). The importance of learning when making inferences. *Judgment and Decision Making Journal, 3*, 261–277.

Rieskamp, J. (in press). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature, 44*, 631–661.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.

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

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

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics, 1*, 43–62.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance, 23*, 86–112.

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

Wallsten, T. S., & Barton, C. (1982). Processing probabilistic multidimensional information for decisions. *Journal of Experimental Psychology: Learning Memory, and Cognition, 8*, 361–384.