

# On the use of recognition in inferential decision making: An overview of the debate

Rüdiger F. Pohl\*

## Abstract

I describe and discuss the sometimes heated controversy surrounding the recognition heuristic (RH) as a model of inferential decision making. After briefly recapitulating the history of the RH up to its current version, I critically evaluate several specific assumptions and predictions of the RH and its surrounding framework: recognition as a memory-based process; the RH as a cognitive process model; proper conditions of testing the RH; measures of using the RH; reasons for not using the RH; the RH as a non-compensatory strategy; evidence for a Less-is-more effect (LIME); and the RH as part of the toolbox. The collection of these controversial issues may help to better understand the debate, to further sharpen the RH theory, and to develop ideas for future research.

Keywords: recognition, recognition heuristic, fast and frugal heuristics, fluency, take-the-best, toolbox, inference, decision making.

## 1 Introduction

As one of the simplest heuristics in the “adaptive toolbox” (Gigerenzer, Todd, & The ABC Research Group, 1999), the recognition heuristic (RH) exploits recognition and may reach a high level of accuracy in inferential decisions. For example, if asked which of two cities is larger, A or B, and given one recognizes A, but not B, one may simply follow the recognition cue and infer that A is the larger city. In domains in which the probability of recognizing an object is substantially related to its criterion value (here the city’s size), such a simple strategy will lead to many correct answers, far above chance. Goldstein and Gigerenzer (1999, 2002) formulated this strategy as the recognition heuristic and defined it as using only one piece of evidence, namely recognition of the two objects (yes/no). No other knowledge about the objects enters the inference process and could possibly overturn the decision based on recognition. The RH thus represents a case of a non-compensatory, one-reason decision-making strategy. Especially this claim has raised some controversy in the past decade and has led to a multitude of new empirical findings. In other words, besides providing a precisely formulated and thus testable

model, one merit of the RH certainly is that it challenged quite a number of researchers, and—as a consequence—extended our knowledge of how inferential decision making may proceed. A new and exciting set of such studies are included in JDMs special issue on “Recognition processes in inferential decision making” (the papers of which can be found in Volume 5, Issue 4, and Volume 6, Issues 1 and 5; see Marewski, Pohl, & Vitouch, 2010, 2011a, 2011b).

In the following section (Section 2), I recapitulate the basic features of the RH and its underlying assumptions, looking at its precursors and its fully laid-out version. In the main part of the paper (Section 3), I then discuss in detail the main points of the controversy surrounding the RH and its framework. Note that I do not try to provide a complete review of all theoretical arguments exchanged so far (see, e.g., Brighton & Gigerenzer, 2011; Bröder & Newell, 2008; Dougherty, Franco-Watkins, & Thomas, 2008; Gigerenzer, 2008; Gigerenzer & Brighton, 2009; Gigerenzer & Gaissmaier, 2011; Gigerenzer & Goldstein, 2011; Hilbig, 2010b, 2011; Hilbig, Erdfelder, & Pohl, 2010; Hilbig & Richter, 2011; Marewski, Gaissmaier, & Gigerenzer, 2010; Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2010; Marewski, Schooler, & Gigerenzer, 2010; Newell & Shanks, 2004; Pachur, Bröder, & Marewski, 2008; Pachur, Todd, Gigerenzer, Schooler, & Goldstein, in press; Tomlinson, Marewski, & Dougherty, 2011; see also the editorial to the first volume of this special issue: Marewski, Pohl, & Vitouch, 2010). Finally (Section 4), I conclude with some general remarks and a short outlook.

The reported research was supported by a grant from the University of Mannheim and by a grant from the DFG (German Science Foundation) to Erdfelder and Pohl (ER 224/2–1). I thank Jon Baron, Arndt Bröder, Benjamin Hilbig, Konstantinos Katsikopoulos, Julian Marewski, and Oliver Vitouch for helpful comments on an earlier version of this paper.

\*Department of Psychology III, University of Mannheim, 68131 Mannheim, Germany. Email: pohl@psychologie.uni-mannheim.de.

## 2 The history of the recognition heuristic

The first ancestor of the RH was mentioned as “familiarity cue” in Gigerenzer, Hoffrage, and Kleinbölting’s (1991) work on probabilistic mental models (PMM). There, in the context of paired comparisons of city names according to the cities’ size, the familiarity cue was defined as “whether one has heard of one city and not the other” (p. 509). This information was considered, for a given domain, as one among five probability cues that govern the building of a PMM and thus the choice behavior (among two alternatives) and corresponding confidence judgments. In an experiment that was planned to test the PMM, the “RH” was born as an explanation for an unexpected finding, namely that the performance of German students who decided which of two cities was larger was about equally good on German and U.S. cities (see Gigerenzer & Goldstein, 2011, and Hoffrage, 2011, for more details on this discovery). Apparently, the German students could exploit recognition (or the lack thereof) to reach such a high performance for the U.S. cities. Accordingly, Gigerenzer and Goldstein (1996)—introducing models of bounded rationality based on the PMM framework—raised the role of recognition information to a “recognition principle” as the first step in their Take-The-Best (TTB) algorithm. The city-size task was raised, too, namely to the status of a “*drosophila*” environment of studying satisficing algorithms (like TTB). The authors assumed that “the recognition principle is invoked when the mere recognition of an object is a predictor of the target variable (e.g., population). The recognition principle states the following: If only one of the two objects is recognized, then choose the recognized object. If neither of the two is recognized, then choose randomly between the two. If both of the objects are recognized then proceed to Step 2.” (p. 653)

Step 2 and further steps then describe how additional cues are searched and evaluated until an inference can be drawn. The authors also stated that the proposed TTB algorithm (including the recognition principle) apply only to inferences from memory (where the cue values have to be retrieved from memory), and not to inferences from givens (where the cue values are openly present to the decision maker).

Following the described precursors, the RH was more fully laid out in Goldstein and Gigerenzer (1999, 2002). It now also received the status of a heuristic on its own. The TTB algorithm was also renamed to a heuristic. RH and TTB and several other heuristics were assumed to form the cognitive tools in an “adaptive toolbox” that human decision makers possess (Gigerenzer et al., 1999). “Adaptive” means that, depending on the task and situation, different, ecologically valid tools could be applied.

Because these tools exploit regularities of the given environment, they allow good and fast decisions with minimal effort. Hence, these strategies were accordingly also termed “fast and frugal heuristics” (FFH; as opposed to more effortful and presumably time-consuming, complex decision processes).

The RH was assumed to be domain-specific, that is, useful only in domains with a high correlation between probability of recognition and criterion value. Recognition was (and is) still used in a binary fashion, that is, objects are either recognized or not. The most important feature, however, was that recognition should be used as the only cue (Gigerenzer & Goldstein, 1996): The authors stated that, if recognition discriminates between the alternatives (i.e., when one object is recognized and the other not), then (1) no other information beyond recognition will be considered and therefore (2) nothing can overturn the inference based on the recognition cue (Goldstein & Gigerenzer, 2002, p. 82). The first hypothesis is known as “one-reason decision making”, the second one as “non-compensatory strategy.” Apparently, these rather bold proposals have fueled the strongest reactions of other researchers (see below).

In addition, the authors introduced the concepts of recognition validity ( $\alpha$ ) and knowledge validity ( $\beta$ ), which can be helpful in describing a domain or a sample. The recognition validity represents the percentage of cases in which following the recognition cue will lead to a correct inference (given that recognition discriminates). The knowledge validity represents the percentage of correct decisions when both objects are recognized (so that recognition does not discriminate). Given that in some domains the recognition validity could excel the knowledge validity, a peculiar effect was predicted, the “less-is-more effect” (LIME). The LIME entails that the overall inferential accuracy of a person who recognizes only about half of the objects in a domain could be higher than that of a person who recognizes all objects. The assumed reason for this at first glance surprising effect is that a person with full recognition can never use the more valid recognition cue, because all objects are recognized and so recognition does not discriminate. Instead, this person has to rely on her (in this case) less valid knowledge. However, a person with fewer recognized objects can utilize the highly valid recognition cue more often and thus will be more often correct.

In both of their central publications on the RH, Goldstein and Gigerenzer (1999, 2002) presented a number of (partly identical, partly different) studies, consisting of experimental work and computer simulations, to support their conjectures as outlined above. These two original publications have sparked a lot of research in the following years up to now, some leading to supporting, others to more critical evidence (to be summarized below).

Table 1: List of controversial topics surrounding the recognition heuristic (RH).

- 
1. Recognition as a memory-based process
  2. The RH as a cognitive process model
  3. Proper conditions of testing the RH
  4. Measures of using the RH
  5. Reasons for not using the RH
  6. The RH as a non-compensatory strategy
  7. Evidence for a Less-is-more effect (LIME)
  8. The RH as part of the toolbox
- 

In their most recent presentation of the RH, Gigerenzer and Goldstein (2011) have clarified the conditions and predictions of the RH theory. The authors also asserted that some of the critical papers that have appeared in the past decade could not be considered adequate tests of the RH (see Section 3.3 and Pachur et al., 2008). Other findings, however, were considered crucial and led to an extension of the RH theory. Most importantly, Gigerenzer and Goldstein now posit that, before the RH is applied, an evaluation will be run that tests whether the recognition cue should be used or not (see Sections 3.5 and 3.8 and Gigerenzer & Brighton, 2009; Marewski, Gaissmaier, Schooler et al., 2010; Pachur & Hertwig, 2006).

### 3 Controversial topics

From looking into the literature (with lots of critical papers, commentaries, and replies), it is clear that the discussion of the “adaptive toolbox” approach and its postulated heuristics has led to a rather lively and sometimes heated debate (see, e.g., the discussion in Gigerenzer & Goldstein, 2011). In this section (and main part of the paper), I present a list of such topics on which researchers diverge or that simply represent open questions to be addressed in the future. I have summarized them under eight headings (Table 1), which I describe and discuss in more detail in the following eight sub-sections.

#### 3.1 Recognition as a memory-based process

While acknowledging that recognition should generally be treated as a continuous variable, Goldstein and Gigerenzer (1999, 2002) focused on the outcome of this recognition process, which is either “recognized” or “not recognized” with only a small and negligible gray zone of uncertainty in between. Accordingly, the quality of these subjective recognition judgments, that is, whether

they were true or not or with what confidence, was originally not considered (see Dougherty et al., 2008, and Newell & Fernandez, 2006, for critical discussions, and Gigerenzer, Hoffrage, & Goldstein, 2008, and Gigerenzer & Goldstein, 2011, for replies). This simplification of the recognition process nevertheless allowed to predict an impressive portion of people’s inferences. Meanwhile, some researchers have asked whether and how the recognition process itself possibly affects subsequent inferences. This question corresponds to Challenge 1 postulated by Tomlinson et al. (2011) and appears even more essential when considering that the proposed heuristics (like RH and TTB) entail memory-search mechanisms, trying to retrieve information (objects, cues, strategies, etc.) from memory. Hoffrage (2011), for example, reported that recognition of city names depended on the size of the reference class from which the cities were drawn, presumably causing a criterion shift for recognition. Another evidence for recognition as continuous variable is that performance for pairs of two “unknown” objects is typically slightly above chance, suggesting that people applied a conservative criterion to “recognize” an object.

One approach that extended the RH theory was presented by Pleskac (2007), who considered recognition in a signal-detection framework (see also Schooler & Hertwig, 2005). He distinguished correctly recognized objects (“hits”) and falsely recognized ones (“false alarms”) and investigated how the proportions of these cases influence the performance of the RH in paired comparisons. Pleskac showed that persons’ sensitivity and their decision criteria affect their performance. Generally, performance of the RH decreases if the number of erroneously recognized objects increases. Another approach, based on a two-high-threshold model of recognition memory, was presented by Erdfelder, Küpper-Tetzl, and Mattern (2011; see also Bröder & Schütz, 2009). According to that model, recognition of items depends on whether the memory strength of “old” objects is above the recognition threshold (leading to “hits”) or not (leading to guessing); and whether the memory strength for “new” objects is below the rejection threshold (leading to “correct rejections”) or not (leading to guessing). Thus, objects could be in a “recognized with certainty” state, in an uncertain state, or in a “unrecognized with certainty” state. Depending on these states and their combinations in pairs of objects, specific predictions about choices and reaction times can be derived. Erdfelder et al. corroborated these predictions in an empirical study, showing the importance of adding a third (uncertain) state to the simple yes/no recognition states used so far.

Another area in which recognition processes are considered concerns the fluency heuristic (FH; Hertwig, Herzog, Schooler, & Reimer, 2008; Schooler & Hertwig,

2005). According to this heuristic, persons use the speed of recognizing an object as another cue. Whenever both objects in a pair are recognized so that the RH cannot be applied, the FH steps in. Given that the fluency of recognition discriminates between the two objects, the FH suggests that the more fluently recognized object should be chosen as having the larger criterion value. This heuristic represents another case of one-reason decision-making. In this context, it is, of course, of paramount interest to understand what determines fluency and how it is perceived and evaluated. In other words, a number of memory search and retrieval processes may play a role here. Hilbig, Erdfelder, and Pohl (2011) estimated the frequency of FH use in cases with both objects recognized and came to a negative conclusion, suggesting that fluency is very rarely considered in isolation as proposed by the FH.

Recently, the applicability of fluency was extended to recognition cases in which only one item is recognized (Marewski, Gaissmaier, Schooler et al., 2010). The authors assumed that the retrieval time for the recognized object determines whether the RH will be applied or not. Whenever retrieval is slow, the decision maker will more likely not use the RH, but will follow it when recognition is fast. Thus, slow retrieval times could be seen as a further reason to stop using the RH (see Section 3.5).

The only sophisticated model so far that took memory processes underlying recognition explicitly into account was presented by Schooler and Hertwig (2005). They implemented RH and FH in the ACT-R cognitive architecture (see, e.g., Anderson & Lebiere, 1998; Anderson & Schooler, 1991) and simulated people's decision processes. In their model, probability of an object's retrieval (for the RH) and its retrieval time (for the FH) are both assumed to be functions of the strength of the object's memory trace and its associative strength to the current retrieval cues. Accordingly, the model allows one to make predictions about whether an object will be recognized, and, if so, how long its retrieval will take. This is certainly an advantage compared to the earlier neglect of memory retrieval processes and also presents a promising test-bed for RH, FH, and other decision processes. However, ACT-R is also a highly complex memory model and necessitates quite a number of assumptions which are not always obvious and which could also be discussed controversially. For example, the specific parameter values can be (and have been) set differently in different model versions, so that empirical predictions were not always that clear.

Another question concerns what it actually means that an object is recognized. Recognition is no doubt helpful in many situations. For example, it helps when someone meets people on the street to know whom to greet (because they are recognized as neighbors) and whom not

(because they are not recognized, suggesting that they are strangers). However, even in this simple situation, it is not recognition itself that is helpful, but rather the information associated with it. Maybe the recognized passerby is someone severely disliked or known for other reasons (because he or she is a famous actor or local politician). In these cases, recognition alone wouldn't suffice to tell what to do. One needs to remember who these persons are, that is, one needs to retrieve further information about them from memory. In other words, it is the combination of recognition and further knowledge that drives behavior in many everyday situations. Newell and Shanks (2004) summarized this by stating that (p. 933) "it is not pure recognition that determines an inference but recognition plus an appropriate reason for knowing why a particular object is recognized—or, at least, a correctly interpreted feeling of familiarity. It is not that an object is recognized and chosen without justification, but that the decision maker has a reasonable idea of why he or she recognizes the object and makes an inference on the basis of this secondary knowledge."

This argument could exemplify why some researchers may feel uneasy that there should be cases in which one's inferences are based on recognition alone. Of course, one may argue that the recognition validity could be low in situations such as the greeting example above (so that the RH would be less useful), but they nevertheless represent cases in which, to be useful, recognition has to be combined with further knowledge. The same argument applies to the classical city-size task, in which cities are not only recognized, but are recognized for being a state's capital, being located at the coast, being a tourist site, or hosting a big automobile company. All this knowledge is intertwined with recognition and is probably retrieved in an instant (see Section 3.6). If that were true, the postulated "search memory" and "stop searching memory" assumptions of the RH possibly need to be changed to inhibitory working-memory processes, trying to prevent any of the already retrieved information beyond recognition to enter the decision making process (see Section 3.5).

That recognition alone could represent an important information can, paradoxically, be shown in cases where recognition is not helpful (Pohl, 2006, Exp. 1). In that experiment, I used a task where recognition was not valid ( $\alpha = .50$ ) and people had (presumably) not much additional knowledge. The task was to decide which of two Swiss cities is located further away from the Swiss city Interlaken, which is close to the geographical center of Switzerland. I found that when one city was recognized and the other not, some participants nearly always inferred that the *recognized* city was the correct one, while another group of participants used exactly the opposite strategy and nearly always chose the *unrecognized* city.



Of course, both groups' accuracy was only around chance (given that recognition was not valid and knowledge not available), but maybe recognition was used as the only "straw" one might cling to, in order to have at least some sense of control in this rather extreme case of decision making. This could be taken as evidence that recognition is indeed an important cue also in other situations.

In sum, it might appear useful to look more closely into the memory processes that lead to the recognition (or rejection) of an object, not just because of extending the RH theory, but rather because these processes presumably have direct consequences on people's behavior and could therefore complement or sharpen predictions as made by the RH alone.

### 3.2 The RH as a cognitive process model

The heuristics in the adaptive toolbox were devised to replace earlier "one-label" or "as-if" models providing more precise descriptions of the processes underlying inferential decision making (see, e.g., Gigerenzer, 1996). As such, some of the postulated heuristics proved quite successful in predicting people's behavior (see, e.g., Gigerenzer & Gaissmaier, 2011). Yet, the next and in my view highly important question is whether and how these heuristics can be translated into cognitive process models, describing how people actually proceed when making an inferential decision (Fiedler, 2010).

Surprisingly, Goldstein and Gigerenzer (1999, 2002) were quite reluctant about using the word "use" in the context of what decision makers are doing with the RH. Of course, the typically reported high adherence rates suggest that the RH is not only understood as a predictive device, but also as an explanation of the processes underlying the observed choices. In addition, the RH has been described in terms of working-memory processes (search, stop, decide) and has accordingly been depicted as a flow chart or production rules (Gigerenzer & Goldstein, 1996; Schooler & Hertwig, 2005; see also Figure 1). Accordingly, Pachur and Hertwig (2006) treated the RH as a cognitive process model and spoke consistently of people *using* the RH or not (see also Pachur et al., in press).

The question then is of how to derive adequate predictions from the RH, for example, for reaction times (RT). I think that the step-wise procedures described in the RH (as well as in other heuristics) should basically allow one to derive such predictions (see, e.g., Glöckner & Bröder, 2011; Hilbig & Pohl, 2009). Moreover, several formulations suggest at least implicit conclusions about RT differences. For example, discussing TTB, Martignon and Hoffrage (1999, p. 137) pointed out that "in the kind of inference task we are concerned with, cues have to be searched for, and the mind operates sequentially, step by

step and cue by cue." Brandstätter, Gigerenzer, and Hertwig (2006) argued with respect to the priority heuristic that it "is intended to model both choice and process: It not only predicts the outcome but also specifies the order of priority, a stopping rule, and a decision rule." (p. 427) In a similar vein, Pachur and Hertwig (2006) claimed that "recognition is first on the mental stage and ready to enter inferential processes when other probabilistic cues still await retrieval." (p. 986) Using recognition should therefore be rather fast, while searching for further information will need additional time (see also Pachur et al., in press).

Supporting evidence for a stepwise TTB process resulting in increasing reaction times the more cues had to be searched was provided by Bröder and Gaissmaier (2007), who had analyzed the data of those participants for which TTB was the best model in predicting choices. Pachur and Hertwig (2006) found that inferences in line with the RH were slower when additional inconsistent information was present.<sup>1</sup> They also reported that under time pressure inferences more often followed the RH. The latter, however, was found in a comparison between different experiments and is therefore difficult to evaluate. Hilbig and Pohl (2009) tested several RT hypotheses that they derived from the RH and contrasted them to an alternative mechanism, namely the difference in evidence (or, in other words, the degree of conflict between the options). In three experiments, they found that most RT results were not compatible with the RH assumptions, but supported the evidence-difference view.

In sum, some more effort should be spent on how to derive predictions for reaction times from the RH, and maybe also for confidence ratings (Glöckner & Bröder, 2011). Having an agreed-upon set of such predictions would help devising experiments, and considering more measures than just choices would better allow to disentangle different explanations.

### 3.3 Proper conditions of testing the RH

Some of the controversy regarding the RH concerned the proper conditions of testing it, and as a consequence, to refuting some of the critical papers as having not followed those conditions (Gigerenzer & Brighton, 2009; Gigerenzer & Goldstein, 2011; Pachur et al., 2008). For example, Pachur et al. (2008) listed eight criteria in which some of the critical studies deviated from the RH theory. These are (1) induced (rather than natural) recognition, (2) induced (rather than natural) cue knowledge, (3) criterion (instead of cue) knowledge, (4) menu-based inferences

<sup>1</sup>That such additional conflicting knowledge was considered by decision makers was also suggested in a study by Hochman, Ayal, and Glöckner (2010) who found that physiological arousal was higher when recognition and additional knowledge were in conflict than when they were not.

(i.e., based on openly given, rather than on memory-retrieved information), (5) domain with low recognition validity, (6) unknown nature of additional cue knowledge, (7) artificial stimuli, and (8) cue knowledge available about unrecognized object.

Firstly, such a list defines and limits the scope of situations where the RH could possibly be tested. This is, on one hand, positive, as it further specifies the RH theory. On the other hand, it restricts the range of potential RH uses to highly specific situations, and thus leads straight to the next question: What does a decision maker do when one, two, or more of these criteria are not met? Are there different heuristics for each of these possible cases? This problem has not been satisfactorily answered yet (see Section 3.8).

Secondly, some of the criteria seem to contradict each other or are at least difficult to control simultaneously. For example, if knowledge may not be learned in the lab, how can the nature of additional cue knowledge be controlled? Nevertheless, Pachur et al. (2008) dismissed the critical findings by Pohl (2006) for exactly that reason, namely that it was unclear what the additional knowledge (which participants in those studies apparently had used) was based on, possibly including some criterion knowledge. Meanwhile, Hilbig, Pohl, and Bröder (2009) have shown that criterion knowledge indeed plays some role (see also Pachur & Hertwig, 2006), but that the main critical findings of Pohl (2006) remain intact when it is controlled for.

Thirdly, Goldstein and Gigerenzer themselves presented a number of studies that did not conform to the list provided by Pachur et al. (2008). For example, Goldstein and Gigerenzer (1999) reported a simulation study and an experiment on (artificial) cue learning. In the experiment, participants could even keep their notes (with the learned cue values) and use these during decision making (as “givens”), so that memory retrieval was not necessary. Goldstein and Gigerenzer (2002) also reported a study in which recognition was experimentally induced by repeatedly testing the same new objects in consecutive sessions (one week apart) which created a sense of (artificial) recognition of these objects in their participants (which does not conform to the criterion of naturally acquired recognition). Or take the case of criterion knowledge. In two of their experiments, Goldstein and Gigerenzer (2002) used sets of German or U.S. cities including the respective largest cities, but did not discuss the potential role of criterion knowledge. Only in a third study was this problem acknowledged and the three largest cities were excluded from the set. In the same paper, the authors wrote (p. 76): “It is also easy to think of instances in which an object may be recognized for having a small criterion value. Yet even in such cases the recognition heuristic still predicts that a recognized object will be

chosen over an unrecognized object.”

This statement directly contradicts the last criterion in Pachur et al.’s (2008) list, but it conforms to Oppenheimer (2003, p. B3) who stated that the RH should be used “even if the recognized city were known to be small.” This prediction (and the corresponding empirical test), despite its equaling Goldstein and Gigerenzer’s consideration, was later criticized as not fulfilling the proper conditions for testing the RH.

All this is, of course, somewhat confusing and may prevent one from “seeing” the proper criteria. One of the main goals of the most recent RH paper by Gigerenzer and Goldstein (2011) was therefore to clarify these conditions. They name three central conditions that define the applicability of the RH: (1) a substantial recognition validity, (2) inferences are made from memory (and not from givens), and (3) recognition stems from natural environments (and not artificial manipulations). Applying this list to published papers would indeed lead to dismiss some of the studies (some with supporting, some with critical findings). Of course, “dismissing” experiments does not imply that these were useless. Rather, they should be seen as testing the boundary conditions of the RH (see Gigerenzer & Goldstein, 2011).

Let me close this section with a short note on one of the relatively undisputed criteria, the recognition validity. The RH is assumed to be useful whenever recognition is a valid cue, but not if it isn’t (see Pohl, 2006, Exp. 1, for supporting evidence). Accordingly, Pachur et al. (2008) dismissed some of the critical studies because recognition validity was apparently low.<sup>2</sup> The underlying and not yet resolved problem, however, is that there is no such thing as an “objective” recognition validity, in the sense that it reflects properties of the real world. The computed validity always depends on two features, namely (a) the set of objects from a domain and (b) the tested participants. For example, if one takes the 20 largest cities of Italy, or the largest 30, or the largest 40, or the 20 cities on ranks 21 to 40, or 41 to 60, or a random sample from all Italian cities with more than 100,000 inhabitants, the resulting recognition validity will differ (see Hoffrage, 2011, for an empirical example). This is why it is important to exactly define the reference class from which the objects are drawn (see, e.g., Gigerenzer & Goldstein, 2011; Pachur et al., in press) and to have a rather large set of such objects to avoid the influence of single, “peculiar” objects. In addition, the recognition validity also

<sup>2</sup>As an aside: They failed to dismiss the results of Pachur and Hertwig (2006) who also had low recognition validities of .60 and .62 in their experiments and thus stated that “recognition is a poor predictor of the criterion in this environment *hostile* to the recognition heuristic.” (p. 989; italics in original) The results were nevertheless cited by Pachur et al. (2008) as supportive for their claim that recognition has a retrieval primacy. Other studies with low recognition validity, but more critical findings, were dismissed.

depends on the sample (e.g., laymen or experts in a domain, or inhabitants from the same country as the cities or from a different one). Given that different persons recognize different objects and different numbers of objects, individual recognition validities will vary. This is fine as long as individual validities are all that is needed. But, to compare data on an aggregate level across experiments, overall recognition validities are necessary. In that case, the mean of individual recognition validities is typically taken as a proxy. But it should be clear that recognition validity represents an abstract concept that is difficult to capture in the real world.

In sum, there has been some debate as to what may count as a proper test of the RH and what rather presents testing its boundaries. The current lists of crucial RH conditions (Gigerenzer & Goldstein, 2011; Pachur et al., 2008) help a lot in this regard and should sharpen future research.

### 3.4 Measures of using the RH

Right from the beginning of doing research on the RH, researchers reported the adherence or accordancy rate, that is, the percentage of times a participant chose the recognized object whenever recognition discriminated. These figures were typically high (90% or higher) and when depicted as individual percentages revealed that a large portion of participants almost *always* chose the recognized object. Respective histogram figures were quite impressive and can be found in many RH papers until today (see Gigerenzer & Brighton, 2009; Goldstein & Gigerenzer, 1999, 2002; Hertwig et al., 2008; Marewski, Gaissmaier, Schooler et al., 2010; Pachur et al., 2008; Pachur & Hertwig, 2006; Reimer & Katsikopoulos, 2004). But what do these rates actually tell us? Bröder and Schiffer (2003) asserted that “simple counting of choices compatible with a model tells us almost nothing about the underlying strategy.” (p. 197) The reason is that we should be careful not to confuse having chosen the recognized object with having applied the RH, because a recognized object may be chosen for a number of reasons, among them recognition. However, other information might have entered the decision process. Typically, retrieved further knowledge about recognized objects correlates positively with the object’s criterion value, that is knowledge is generally confounded with recognition. Thus, without further measures one cannot tell whether inferences were based on recognition, on other knowledge, on recognition plus other knowledge, or on guessing (Hilbig, 2010a). Hilbig (2010b) demonstrated this obvious, but often neglected fallacy in a convincing way by introducing a nonsense heuristic which nevertheless “explained” a significant proportion of choices. Tomlinson et al. (2011) have addressed this problem as one of their main challenges to

the current RH research.

Moreover, which role does guessing play? We would assume “guessing” if an adherence rate was 50%, because there was no clear tendency to choose the recognized object more often than the unrecognized one. Let us assume we have an adherence rate significantly above chance, say 70%. Does that mean that the RH was (potentially) followed in 70% of the cases? Probably not, because the remaining percentage (of  $100 - 70 = 30\%$ ) is likely due to guessing processes (assuming that nothing spoke explicitly against choosing the recognized object), and therefore the same portion of choices conforming to the RH would presumably also have resulted from guessing (30%). Only the remaining ( $100 - 30 - 30 = 40\%$ ) might be indicative of the RH, which is less impressive than the adherence rate of 70%. Of course, if adherence rates are as high as typically reported, guessing apparently plays only a minor role.

Apart from guessing: What happened when someone chose the unrecognized object? Was knowledge involved that spoke *against* the recognized object? Again, from adherence, or non-adherence, rates we cannot tell. We need further information to understand what actually caused the observed choice behavior. Therefore, other measures that went beyond simple adherence rates were introduced, namely a discriminability parameter based on signal detection theory ( $d'$ ; Pachur & Hertwig, 2006) and a discrimination index (DI; Hilbig & Pohl, 2008). Pachur and Hertwig focused on how well participants can discriminate between the recognized object representing a correct or false inference. A correctly chosen recognized object would then represent a hit, a falsely chosen recognized one a false alarm. From these proportions, they computed  $d'$  as an estimate of a participant’s discrimination ability. This index should be zero if only recognition was used. But it wasn’t, suggesting that participants were to some extent able to distinguish between valid and invalid RH-based inferences. In a similar vein, the DI computes how often the recognized object was chosen when it was in fact the correct choice, minus the number of choices when it was the false one. This index should be zero, if participants use only recognition and can thus not discriminate between recognized objects being correct or false. If the index is different from zero (as Hilbig & Pohl, 2008, consistently found), some further information in addition to or instead of recognition must have been used. When applied to individual data, the DI suggested that the majority of participants did not use the RH.

In a recent attempt to overcome the problems of adherence rates, Hilbig, Erdfelder, and Pohl (2010) proposed and validated a multinomial processing tree model, named the r-model, as a measurement tool to yield bias-free estimates for the probability of RH use. Their

general result was that these estimates are significantly smaller than adherence rates suggest, but still significantly above chance (see also Hilbig et al., 2011, for an extension of the *r*-model to measure use of the FH). Hilbig (2010a) compared the different measures of RH use (adherence,  $d'$ , DI, and  $r$ ) in simulation studies and found that the *r*-model delivered the best results.<sup>3</sup> But note that the *r*-model is simply a measurement tool and not a theoretical model, that is, it does not explain why people did or did not use the RH in their inferences. It only estimates the respective frequencies.

In sum, while accordance rates as a measure of RH use appear faulty because they are confounded, other measures have been introduced that allow better estimates of how often the RH was used. The *r*-model provides the latest of these measures and could prove a helpful tool in testing the RH.

### 3.5 Reasons for not using the RH

One argument that could be used to explain evidence that is contradictory to the RH is to assume that people decide in each case whether the RH would be the best strategy to apply. If not, they use some other strategy. Pachur and Hertwig (2006, p. 993) stated that “people appear to decide case by case whether they will obey the recognition heuristic. Moreover, these decisions are not made arbitrarily but demonstrate some ability to discriminate between cases in which the recognition heuristic would have yielded correct judgments and cases in which the recognition heuristic would have led astray.”

They also assumed that the RH is typically chosen as the default strategy in recognition cases (i.e., whenever one object is recognized and the other not), but that it can be “suspended” for a number of reasons and thus not applied to the current case. The reasons for suspending the RH include (1) availability of probabilistic cues with larger validities than the recognition validity; (2) source knowledge (i.e., knowing that an object is recognized for other reasons than its criterion value; e.g., Chernobyl is recognized by most people, but not because of its size, but because of the nuclear accident in 1986); and (3) conclusive criterion knowledge. These reasons could explain why the RH is not applied in every single case.

The third reason is probably the most obvious one. If criterion knowledge is available, that is, knowledge that allows a direct conclusion whether or not the recognized city is small or large, the decision (for or against the recognized city) can be directly deduced from the available knowledge. A probabilistic inference such as the RH will then be superfluous.<sup>4</sup> But the problem for

this and the first two potential reasons for suspending the RH is conceptual: Before the RH can be applied, all available knowledge needs to be retrieved and scanned whether it contains anything that speaks against applying the RH. Thus, memory search cannot stop as soon as recognition is assessed as the RH assumed. Accordingly, Pachur and Hertwig (2006) suggested a two-stage-process, in which recognition is followed by an evaluative step that determines whether the RH should be applied (see also Gigerenzer & Brighton, 2009, p. 132; Gigerenzer & Goldstein, 2011; Marewski, Gaissmaier, Schooler et al., 2010).

Just recently, Marewski, Gaissmaier, Schooler et al. (2010) added another reason why the RH might not get applied. They assumed that the retrieval time (i.e., the time to decide whether an object is recognized or not) can be used as a cue: When the recognized object is retrieved fast, persons should go more often with the RH than when it is retrieved slowly. The rationale for this is that objects with further available cue knowledge are typically retrieved (recognized) faster than objects without additional knowledge. And since additional knowledge more often speaks *for* the recognized object than *against* it, it would be wise to go with the RH. In other words, faster recognition times (fluency) could simply be taken as a proxy for the existence of additional information that speaks for the recognized object. A slow retrieval, however, would signal that no additional information is available that would possibly speak for the recognized object. In this case, one should hesitate to go with recognition and thus not use the RH.<sup>5</sup>

Two of the given reasons for suspending the RH have an important implication. If an inference is based on a more valid knowledge cue or on a slow recognition time, leading to suspending the RH, this inference may nevertheless choose the recognized object. It is thus clear that simple adherence rates generally overestimate use of the RH and that it depends on the proportions of these other cases as to how much its use is overestimated (see Section 3.4 and Gigerenzer & Goldstein, 2011; Hilbig, 2010b; Hilbig, Erdfelder, & Pohl, 2010; Hilbig & Pohl, 2008).

In sum, the decision process is more complicated than previously assumed. Besides, proposing that, before the RH is applied, memory is searched in order to check whether anything better can be found or whether something speaks against using the RH, appears tantamount to saying that recognition is used as a cue whenever nothing

<sup>3</sup>*k* objects in a set allows one to deduct the answer in all recognition cases involving one of the *k* objects.

<sup>5</sup>One apparent problem with this conception is that the validity of recognition depends on retrieval fluency which in turn depends on the availability of knowledge that speaks for the recognized object. In other words, recognition validity would not only depend on recognition but also on further knowledge.

<sup>3</sup>The software to run the *r*-model (multiTree; Moshagen, 2010) and the appropriate equation files are available from the authors.

<sup>4</sup>For example, criterion knowledge consisting of knowing the largest



better is available. But then the RH is no longer a shortcut, intentionally ignoring other, potentially useful information. Moreover, similar arguments may apply to the FH, where it needs to be checked whether fluency can be used as a cue or whether it should be attributed to some other source (not related to the criterion) and therefore discarded.

### 3.6 The RH as a non-compensatory strategy

As I have discussed above (in Section 3.4), choosing the recognized object is not identical with basing one's decision on recognition alone (see Hilbig, 2010a, 2010b). In natural environments, recognition and knowledge are most likely confounded, that is, the probability of recognition increases with the cities' size and so does cue knowledge that speaks for the cities' largeness. Pohl (2006), for example, reported differences in adherence rates, depending on (a) whether participants merely recognized the object's name or knew more about it, and (b) whether the recognized object was actually the correct choice or not. When participants knew more about the recognized object or when it represented the correct choice, they chose it consistently and significantly more often (see also Hilbig & Pohl, 2008; Newell & Fernandez, 2006; Oeusoonthornwattana & Shanks, 2010; Pachur et al., 2008; Richter & Späth, 2006). These results suggest that more (or something different) than recognition was used in these inferences and it remains unclear how often (or whether at all) an inference was based on recognition alone. Note that these observations could be based on two cases: (1) The recognized object was chosen for another reason than recognition alone; and (2) the unrecognized object was chosen despite non-recognition. The latter could possibly represent compensatory inferences, where the decision based on recognition was overturned by another cue (that spoke for the unrecognized or against the recognized object), thus suggesting that something different than the non-compensatory RH was used in these cases. But they could also result from yet some other non-compensatory mechanism not considering recognition at all. One approach to tackle these problems is to formulate and test respective compensatory and non-compensatory models (see Marewski, Gaissmaier, Schooler, et al., 2010).

In other cases, even if evidence that is contradictory to the recognition cue is retrieved and considered, the impact of this evidence might be too weak to overturn the inference based on recognition. This could be a matter of (subjective) validity. If recognition has a high validity and the additional knowledge a low one, this additional knowledge will most likely fail to dominate the decision. So, only if the additional cue's validity is large enough, it

could eventually overrule recognition. Similarly, Pachur and Hertwig (2006) argued that one possible reason for not using the RH would be if an additional probabilistic cue had a larger validity than the recognition information (see also Newell & Shanks, 2004). In sum, given that the RH applies only in situations with large recognition validities, situations in which the validity of additional knowledge is even larger could be rare. Accordingly, most decisions in such domains might indeed be non-compensatory based on recognition only.

One way to test this assumption is to experimentally control participants' cue knowledge by introducing additional knowledge cues which validly contradict the recognition cue and then to observe the according choices (Goldstein & Gigerenzer, 2002; Newell & Fernandez, 2006; Pachur et al., 2008; Richter & Späth, 2006). This procedure, however, would not conform to the "proper" criteria as defined above and might not get accepted as a test of the RH (see Section 3.3). A summary of such research is given in Pachur et al. (in press). The evidence suggests that, on an aggregate level, mean adherence rates drop somewhat when additional, contradictory evidence is present, but that the effect is much smaller when analyzed on an individual level, showing that a large portion of participants chooses the recognized object irrespective of any contradictory evidence. Only some participants apparently change their strategy.<sup>6</sup>

### 3.7 Evidence for a Less-is-more effect (LIME)

The LIME is defined as a pattern of results in which recognition of fewer objects leads to more accurate inferences than recognition of more objects does. One question concerns the conditions under which such an effect is predicted. Goldstein and Gigerenzer (1999, 2002) argued that (1) the recognition validity  $\alpha$  must be higher than the knowledge validity  $\beta$  and that (2)  $\alpha$  and  $\beta$  remain constant across the number of recognized objects ( $n$ ). But, after presenting a simulation study, they added (Goldstein & Gigerenzer, 2002), that "the simplifying assumption that the recognition validity  $\alpha$  and knowledge validity  $\beta$  remain constant is not necessary for the less-is-more effect to arise." (p. 81)

The classical example of the LIME (with three Scottish brothers, or Parisian sisters; Goldstein & Gigerenzer, 1999, 2002) was unfortunately not too enlightening in this respect. The authors assumed that Brother A recognizes none of the, say 20, objects ( $n = 0$ ), Brother B recognizes half ( $n = 10$ ), and Brother C all ( $n = 20$ ). Now consider the  $\alpha$  and  $\beta$  values of these three persons. Brother A, recognizing no object, has to guess all the

<sup>6</sup>However, as discussed above, one problem is that these adherence rates could be flawed (Hilbig, 2010a, 2010b; Hilbig & Richter, 2011).

time, that is, he has neither a recognition nor a knowledge validity (both are not defined in this case). For Brother B the authors assume, for example, a recognition validity of .80 and a (lower) knowledge validity of .60. For Brother C, the recognition validity cannot be determined, because he recognizes all objects. His knowledge validity is also assumed to be .60. In sum, only one brother has a recognition validity (nothing can be known about the other two), and two brothers are assumed to have the same knowledge validity (nothing can be known about the third one). Thus it remains unclear from these examples, too, how  $\alpha$  and  $\beta$  actually behave or should behave relative to  $n$  (see Dougherty et al., 2008, for a discussion of further critical details, and Gigerenzer et al., 2008, for a reply).

Meanwhile, a number of studies extended the originally formulated conditions. One finding is that people's memory sensitivity to distinguish between recognized and non-recognized objects should be high (Pleskac, 2007; see Section 3.1), another that decision makers should actually behave as the RH assumes (see Hilbig et al., 2009). Pachur (2010) tested the above mentioned validity dependencies (i.e., the correlations between  $n$  and both validities  $\alpha$  and  $\beta$ ) in computer simulations and found that they could have a strong limiting effect on the LIME. Katsikopoulos (2010) showed that the relation  $\alpha > \beta$  is not a necessary precondition (see also Beaman, Smith, Frosch, & McCloy, 2010; Davis-Stober, Dana, & Budescu, 2010; and Smithson, 2010; for still other variants of the LIME). Thus, there appear to be several situations in which a LIME may occur. Theoretically, the LIME could be of quite a large size. Assuming strict adherence to the RH and extreme values, namely a recognition validity of 1.0 and a knowledge validity of .50, the effect reaches its maximum with a difference of 26.3% (Pohl, 2006), that is, a person recognizing all objects will show a percentage of correct inferences that is 26.3% below the performance of someone who recognizes less, but just the right number of objects so that the high recognition validity can be most effective (in this case, the optimal number would be to recognize 50% of the objects).

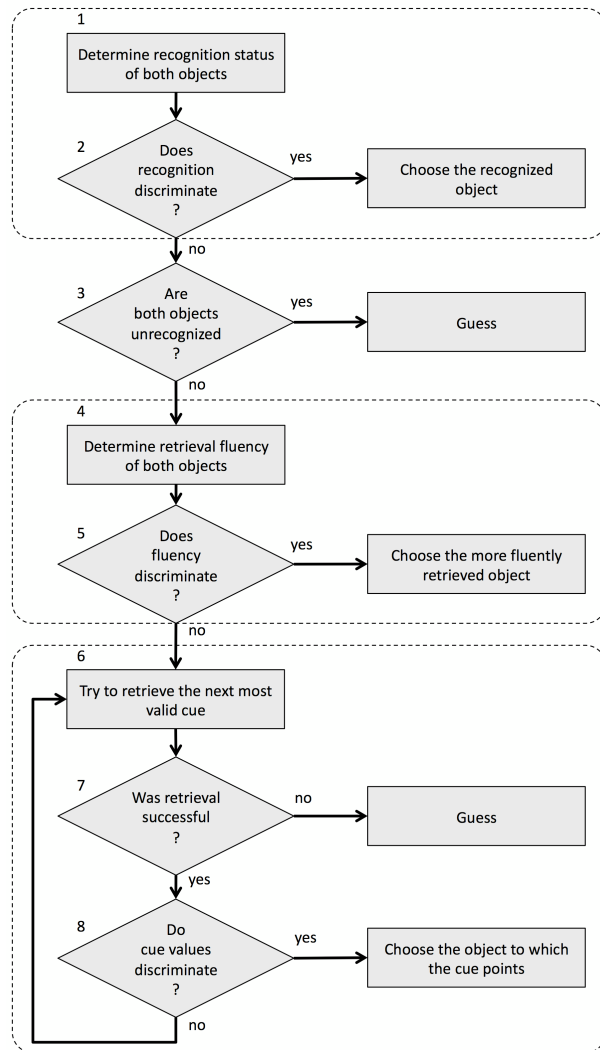
Another question is whether the LIME has been shown empirically so far. Most of the manifestations are based on simulations only. Goldstein and Gigerenzer (2002) thus admitted that "the curious phenomenon of a less-is-more effect is harder to demonstrate with real people than by mathematical proof or computer simulation." (p. 83) In one of their studies, they interpreted a performance difference of 0.3% as showing a slight LIME (without reporting a statistical test). In a second study, they used experimentally induced recognition and found a significant LIME of 3.5%. However, that induction procedure was later criticized by themselves as well as by Pachur et al. (2008) as not conforming to the proper RH con-

ditions (see also Marewski, Gaissmaier, Schooler et al., 2010; and Section 3.3). Pohl (2006) computed theoretically possible LIMEs in eight data sets and found that the LIME was not predicted in four sets (because  $\alpha \leq \beta$ ) and rather small in the remaining sets (ranging from 2.2 to 8.0%). Computing the real LIME was unfortunately not possible, because the range of recognized objects was too small (but see Pachur, 2010, who computed predicted accuracy curves for those data). In Exp. 3, Pohl (2006) compared different domains (namely Belgian, Italian and German cities). Participants had mean recognition rates of 6.6, 9.5, and 11.0 (out of 11 cities each) for these three domains, yet performance increased significantly with the number of recognized cities, that is, it showed a "more-is-more" effect (see also Pachur & Biele, 2007).

Using a design in which inferences were recorded from groups rather than individual persons, Reimer and Katsikopoulos (2004) reported cases of LIMEs ranging from 2 to 8%, but without reporting statistical tests. Besides, they used a rather lax criterion to define a LIME. Whenever there exist two persons (or in this case, groups) with different numbers of recognized objects,  $n_1$  and  $n_2$ , such that  $n_1 < n_2$ , then a LIME is said to occur if the performance is higher for  $n_1$  than for  $n_2$  (see also Pachur, 2010). The problem is that such cases simply *must* occur just by chance (unless individual performance data are perfectly monotonically ordered along values of  $n$ ). For example, Pachur et al. (in press) cited results from Snook and Cullen (2006) as showing a LIME, but they had picked two single participants out of the sample (see Fig. 5 of Snook & Cullen, 2006), with one participant having the highest percentage of correct inferences (86%) and recognizing about half the objects, and the other one recognizing the most objects, but performing less well (76%). Hence, these two persons "show" a LIME. Such selective comparisons appear questionable as long as they are not guarded against chance results. The Snook and Cullen (2006; Figure 5) data nicely demonstrate this problem as it is easy to find pairs of persons with the opposite pattern. For example, when one picks the two persons with the highest number of recognized objects, they show a clear "more-is-more" effect. In a further analysis, Pachur (2010) again used the data from Snook and Cullen (2006), but ran a regression analysis. The results suggested a quadratic relationship between the number of recognized objects and accuracy, which would be indicative of a LIME (but see Figures 2 and 5 of Pachur, 2010).<sup>7</sup>

<sup>7</sup>Note that Snook and Cullen (2006) did not interpret their own data as showing a LIME: "As the total number of players participants recognized [...] increased, there was a corresponding increase in accuracy. [...] this trend was maintained until approximately half of the 200 players were recognized. When the recognition rate was above approximately half of the players, accuracy leveled off." (p. 40) While this may be seen as a valid description of their results, one problem with these data remains: Only a few participants recognized more than half

Figure 1: Schematic flow-chart representation of different heuristics and decision processes that are possibly involved in paired comparisons; the areas surrounded by dashed lines represent (from top to bottom) the RH, the FH, and TTB (not shown are additional evaluations of whether the RH, the FH, or TTB should be used in a given situation).



In sum, several recent studies have more deeply explored the conditions under which a LIME could theoretically be expected, thus extending earlier formulations of this phenomenon. The empirical evidence for a LIME, however, remains scarce with the reported effects mostly being of minor size. Perhaps it is difficult to find real domains that exactly possess those conditions that theoretically foster a LIME.

of the players. It is therefore rather difficult to decide which function fits the data best (see Pohl, 2006, for a similar problem).

### 3.8 The RH as part of the toolbox

In typical experiments using paired comparisons (e.g., with city names), participants answer a series of such comparisons and infer which of the two objects in each pair is the larger one. For example, in a set of 20 objects and with all possible pairwise combinations, participants work through some 190 trials. Given that not all objects are recognized or not all are unrecognized, there will be different types of pairs, or “cases”, depending on how many of the objects in a pair are recognized: (1) Recognition cases, in which one object known and the other not. These cases represent the central ones for studying the RH. (2) Guessing cases, consisting of two unknown objects, such that persons have nothing left but to guess (or to infer probabilistic cues from the names of the objects, e.g., to which country a city might belong, thus allowing inferences about its size). Recognition is not helpful here, because none of the objects is recognized. (3) Knowledge cases, consisting of pairs in which both objects are known. Again, recognition is of no help since both are recognized. In this case, other knowledge has to be assessed to reach a decision. This could be the fluency of retrieving the objects from memory, such that persons infer that the faster retrieved object is possibly the larger one (FH; Schooler & Hertwig, 2005; Hertwig et al., 2008; but see Hilbig et al., 2011, for conflicting findings). Or, if fluency is similar for both objects, further cue knowledge must be invoked. Here, still another heuristic, namely Take-the-Best (TTB), comes into play. According to TTB, knowledge cues are searched one by one following their cue validity. As soon as one cue discriminates between the two objects, search will stop and the decision will be made based on that cue. Again, further knowledge is ignored. If all fails, one must guess.

This cascaded decision tree is depicted as a flow chart in Figure 1 (see Gigerenzer & Goldstein, 1996, Fig. 2; Schooler & Hertwig, 2005, Tables 1 and 2; for similar descriptions). The areas surrounded by dashed lines represent the three different heuristics involved: the upper area includes the RH, the middle area the FH, and the bottom area the TTB heuristic. Note that this chart is only meant to be a summary of all potential decision steps, and not a strictly serial process model of how a decision maker actually proceeds. The chart nevertheless shows that the studied paired comparisons are more complicated than each of the “fast and frugal” heuristics when viewed as a single strategy suggests.

Even more complicating, according to the recently proposed evaluation stage, memory has to be searched in every single recognition case whether any information is available that would argue against using the RH (see Section 3.5; Gigerenzer & Goldstein, 2011; Marewski, Gaissmaier, Schooler et al., 2010; Pachur & Hertwig,

2006). These evaluative processes are not shown in Figure 1, but they—together with the other two heuristics (FH and TTB), for which similar evaluative processes may hold (see Marewski, Gaissmaier, & Gigerenzer, 2010)—make it questionable whether deciding which of two objects is larger still appears as “fast and frugal” as the FFH approach originally assumed.

Inherent in this description is also one of the main problems of the toolbox approach (see Glöckner, Betsch, & Schindler, 2010; Newell, 2005; Newell & Lee, 2010): How does one know when to take which heuristic? This corresponds to Challenge 3 postulated by Tomlinson et al. (2011) and also to one of the five main research questions posited by Marewski, Schooler, and Gigerenzer (2010). Goldstein and Gigerenzer (2002; see also Gigerenzer & Gaissmaier, 2011) suggested that the knowledge which strategy should be applied in which situation might be either (a) genetically coded, (b) socially or culturally transmitted, or (c) learned individually (see also Rieskamp & Otto, 2006). While these mechanisms seem plausible, they also remain somewhat vague yet, so that the strategy-selection problem is certainly an area in which more research is needed.

The heuristic selection in experimental studies is further complicated by the fact that this decision has to be made anew for each of the, say, 190 trials (given 20 objects and all combinations), in which the different types of pairs appear in random order. One cannot a priori stick to the same heuristic for the next trial. This makes the repeated traversing through some or all potential decision steps (as depicted in Figure 1) look quite strenuous.<sup>8</sup>

## 4 Conclusions

In this paper, I started with a short description of the development of the theory underlying the recognition heuristic (RH) and then discussed at length some of its controversial issues. Note that the selection of these issues and their handling reflects my personal preferences and opinions. As such, this paper was not intended to be “neutral”, although I nevertheless strove for a (sometimes more, sometimes less) balanced presentation. Others, no doubt, would have focused on other topics and would presumably have come to other conclusions (see, e.g., Gigerenzer & Brighton, 2009; Gigerenzer & Gaissmaier, 2011; Gigerenzer & Goldstein, 2011; Marewski, Gaissmaier, & Gigerenzer, 2010; Marewski, Gaissmaier, Schooler et al., 2010; Marewski, Schooler, and Gigerenzer, 2010; Tomlinson et al., 2011). What I think all

<sup>8</sup>One might object that such situations of repeated judgments in paired comparisons are not very common in real life. The typical experimental “drosophila” procedure might therefore not be ecologically valid, and thus there is no need to assume that mental toolboxes are adapted to it. But then, we should start looking for more valid tasks.

involved researchers would agree upon is that the RH and the FFH framework represent an enormous advantage over previous conceptions. The precise formulations and sometimes bold predictions, moreover, fueled not only the described debate but also a wealth of empirical research, leading to the development of new methods and theoretical ideas. For example, the RH was extended (a) from single participant’s inferences to a “wisdom-of-crowd” measure (Gaissmaier & Marewski, 2011; Herzog & Hertwig, 2011); (b) from paired comparisons to multi-alternative decisions (Frosch, Beaman, & McCloy, 2007; Marewski, Gaissmaier, Schooler et al., 2010; McCloy, Beaman, & Smith, 2008); or (c) from inferences to preferences (Oeusoonthornwattana & Shanks, 2010). As such, the whole field has certainly benefited.

Let me summarize the empirical findings on the RH with a quote from Pachur et al. (2008, p. 205) who stated that “it is now clear that the recognition heuristic—in particular in terms of the hypothesized non-compensatory use of recognition—is not used by all people all the time and under all circumstances.” And that (p. 206) “individuals appear to differ greatly in their reliance on recognition for inferences.” These conclusions may also lead the way to future research, that is, to further define the influences of domains, tasks, and individual characteristics on which strategy is preferred in which situation. Thus, one viable and legitimate question, asked from the early days of the FHH approach, is which heuristics like the RH are suited for which domains and tasks. This certainly helps to define the boundary conditions of the RH and any other heuristic. Another question, although also asked from the beginning, namely that of individual differences, might be more difficult to answer. First of all, if the “adaptive” use of such decision strategies like the RH reflects environmental regularities, why should individuals differ so much in their perception or evaluation of these regularities? Why should, within the same domain, some people rely on recognition almost always, and others only occasionally (as has been reported in some studies)? This in my view is still somewhat puzzling, although some preliminary and tentative answers regarding individual difference in use of heuristics have meanwhile appeared (e.g., Bröder, 2003; Hilbig, 2008; Pachur, Mata, & Schooler, 2009).

In sum, the general question concerning the RH would then not be to ask whether it is used but rather when and by whom it is used (see, e.g., Gigerenzer & Brighton, 2009; Hilbig, Erdfelder, & Pohl, 2010; Hilbig, Scholl, & Pohl, 2010; Pachur & Hertwig, 2006; Pohl, 2006). When phrased in such a way, the current controversy surrounding the RH loses much of its impetus and one may wonder why such a simple question has raised so many debates. One answer could be that some have not stopped there and have instead questioned the RH as a valid tool



and finally the whole FFH approach (see Dougherty et al., 2008; Fiedler, 2010; Glöckner & Betsch, 2008a; Glöckner et al., 2010; Hilbig, 2010b; Newell, 2005). One reason for such a fundamental critique may be grounded in the FFH's central assumption that there are a number of different tools available from which the decision maker has to choose the appropriate one that best fits a given environment. Moreover, according to the theory, once a potentially useful strategy is identified, the decision maker has to check whether any reason would speak against using that otherwise optimal tool. Pachur and Hertwig (2006) listed a number of such reasons why the RH could be "suspended" (see Section 3.5; Gigerenzer & Goldstein, 2011). All these reasons may lead to the selection, evaluation, and (finally) application of other tools. Such a series of strategy selection and evaluation steps, repeatedly for each single trial, seems quite cumbersome (see Section 3.8) and too complicated to work in practice. This is especially so, if numerous paired comparisons have to be made in a row (e.g., 190 different ones for a set of 20 objects), as is typical for the experiments that have been run to study these heuristics.

In addition to this more theoretical argument, some researchers came to a negative conclusion regarding the FFH approach when summarizing the available empirical evidence.

For example, Hilbig (2010b, p. 923) concluded that "the empirical evidence available does not warrant the conclusion that heuristics are pervasively used." Similarly, Fiedler (2010, p. 22) asserted that "it seems fair to conclude that strict empirical tests have resulted in a more critical picture of the validity and scope of the postulated heuristics."

In this situation, alternative conceptions that posit fewer or only one mechanism instead of multiple tools have been proposed and may gain ground (see Hilbig & Pohl, 2009). Without going into too much detail I mention only two, namely the evidence-accumulation models, reappearing in the "adjustable-spanner" metaphor (Lee & Cummins, 2004; Newell, 2005; Newell, Collins, & Lee, 2007; Newell & Lee, 2010), and the recently proposed "Parallel Constraint Satisfaction" network model (PCS; Glöckner & Betsch, 2008a, 2008b; Glöckner et al., 2010; Glöckner & Bröder, 2011). One of the major advantages of these approaches is that they can be applied to all comparison types (not just recognition cases) and also easily combine compensatory and non-compensatory use of probabilistic cues within the same architecture and thereby avoid the need to change tools from one trial to the next. For example, Glöckner and Bröder (2011) tested the RH against the PCS, albeit in a different situation than the RH was proposed for, namely with cue values openly available to the participants (as "givens") and also for unrecognized alternatives. Using a maximum-likelihood

classification method (including choices, response times, and confidence ratings) the authors found that 77.5% of their participants' behavior was best explained by the PCS strategy and that only a small portion of participants (up to 7.5%) were classified as RH users. Newell and Lee (2010) also used a "givens"-procedure and tested a sequential evidence-accumulation approach (SEQ) against TTB. Using a minimum-description-length criterion (to account for the different complexities of the models), they reported that the pattern of results was best captured by their SEQ model treating TTB as a special subcase. Comparing these alternative models to the toolbox approach then really is a bigger controversy (than just discussing the rate of RH use). Of course, it remains to be seen how these alternatives succeed in the originally proposed inferences-from-memory situation (but see Hilbig & Pohl, 2009). Thus it is still too early to draw any further conclusions about how good these alternatives will fare in the end. But it is quite clear that there is more, perhaps even more fundamental, debate to come in the near future (see, e.g., Glöckner & Betsch, 2010; Marewski, 2010).

## References

- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- Anderson, J. R., & Schooler, J. W. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- Beaman, C. P., Smith, P. T., Frosch, C. A., & McCloy, R. (2010). Less-is-more effects without the recognition heuristic. *Judgment and Decision Making*, 5, 258–271.
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113, 409–432.
- Brighton, H., & Gigerenzer, G. (2011). Towards competitive instead of biased testing of heuristics: A reply to Hilbig and Richter (2011). *Topics in Cognitive Science*, 3, 197–205.
- 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 multiattribute decisions. *Psychonomic Bulletin & Review*, 14, 895–900.
- Bröder, A., & Newell, B. R. (2008). Challenging some common beliefs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making*, 3, 205–214.
- Bröder, A., & Schiffer, S. (2003). Take The Best versus simultaneous feature matching: Probabilistic infer-

- ences from memory and effects of representation format. *Journal of Experimental Psychology: General*, *132*, 277–293.
- Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 587–606.
- Davis-Stober, C. P., Dana, J., & Budescu, D. (2010). Why recognition is rational: Optimality results on single-variable decision rules. *Judgment and Decision Making*, *5*, 216–229.
- Dougherty, M. R. P., 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*, 199–213.
- Erdfelder, E., Küpper-Tetzl, C. E., & Mattern, S. D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making*, *6*, 7–22.
- Fiedler, K. (2010). How to study cognitive decision algorithms: The case of the priority heuristic. *Judgment and Decision Making*, *5*, 1–12.
- Frosch, C. A., Beaman, C. P., & McCloy, R. (2007). A little learning is a dangerous thing: An experimental demonstration of recognition-driven inference. *The Quarterly Journal of Experimental Psychology*, *60*, 1329–1336.
- Gaissmaier, W., & Marewski, J. N. (2011). Forecasting elections with mere recognition from small, lousy samples: A comparison of collective recognition, wisdom of crowds, and representative polls. *Judgment and Decision Making*, *6*, 73–88.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A rebuttal to Kahneman and Tversky (1996). *Psychological Review*, *103*, 592–596.
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*, 20–29.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristics: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, *6*, 100–121.
- Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Postscript: Fast and frugal heuristics. *Psychological Review*, *115*, 238–239.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
- Gigerenzer, G., Todd, P. M., & The ABC Research Group (Eds.). (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Glöckner, A., & Betsch, T. (2008a). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making*, *3*, 215–228.
- Glöckner, A., & Betsch, T. (2008b). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1055–1075.
- Glöckner, A., & Betsch, T. (2010). Accounting for critical evidence while being precise and avoiding the strategy selection problem in a parallel constraint satisfaction approach: A reply to Marewski (2010). *Journal of Behavioral Decision Making*, *23*, 468–472.
- Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*, *23*, 439–462.
- Glöckner, A., & Bröder, A. (2011). Processing of recognition information and additional cues: A model-based analysis of choice, confidence, and response time. *Judgment and Decision Making*, *6*, 23–42.
- Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 37–58). New York: Oxford University Press.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.
- Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1191–1206.
- Herzog, S. M., & Hertwig, R. (2011). The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgment and Decision Making*, *6*, 58–72.
- Hilbig, B. E. (2008). Individual differences in fast-and-frugal decision making: Neuroticism and the recognition heuristic. *Journal of Research in Personality*, *42*, 1641–1645.
- Hilbig, B. E. (2010a). Precise models deserve precise measures: A methodological dissection. *Judgment and Decision Making*, *5*, 272–284.
- Hilbig, B. E. (2010b). Reconsidering “evidence” for fast and frugal heuristics. *Psychonomic Bulletin & Review*, *17*, 923–930.

- Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2010). One-reason decision-making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 123–134.
- Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2011). Fluent, fast, and frugal? A formal model evaluation of the interplay between memory, fluency, and comparative judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 827–839.
- Hilbig, B. E., & Pohl, R. F. (2008). Recognizing users of the recognition heuristic. *Experimental Psychology*, *55*, 394–401.
- Hilbig, B. E., & Pohl, R. F. (2009). Ignorance- versus evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1296–1305.
- Hilbig, B. E., Pohl, R. F., & Bröder, A. (2009). Criterion knowledge: A moderator of using the recognition heuristic? *Journal of Behavioral Decision Making*, *22*, 510–522.
- Hilbig, B. E., & Richter, T. (2011). Homo heuristicus outnumbered: Comment on Gigerenzer and Brighton (2009). *Topics in Cognitive Science*, *3*, 187–196.
- Hilbig, B. E., Scholl, S. G., & Pohl, R. F. (2010). Think or blink: Is the recognition heuristic an “intuitive” strategy? *Judgment and Decision Making*, *5*, 300–309.
- Hochman, G., Ayal, S., & Glöckner, A. (2010). Physiological arousal in processing recognition information: Ignoring or integrating cognitive cues? *Judgment and Decision Making*, *5*, 285–299.
- Hoffrage, U. (2011). Recognition judgments and the performance of the recognition heuristic depend on the size of the reference class. *Judgment and Decision Making*, *6*, 43–57.
- Katsikopoulos, K. V. (2010). The less-is-more effect: Predictions and tests. *Judgment and Decision Making*, *5*, 244–257.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the “take the best” and the “rational” models. *Psychonomic Bulletin & Review*, *11*, 343–352.
- Marewski, J. (2010). On the theoretical precision and strategy selection problem of a single-strategy approach: A comment on Glöckner, Betsch, and Schindler (2010). *Journal of Behavioral Decision Making*, *23*, 463–467.
- Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing*, *11*, 103–121.
- Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multialternative inference. *Psychonomic Bulletin & Review*, *17*, 287–309.
- Marewski, J. N., Pohl, R. F., & Vitouch, O. (2010). Recognition-based judgments and decisions: Introduction to the special issue (Vol. 1). *Judgment and Decision Making*, *5*, 207–215.
- Marewski, J. N., Pohl, R. F., & Vitouch, O. (2011a). Recognition-based judgments and decisions: Introduction to the special issue (II). *Judgment and Decision Making*, *6*, 1–6.
- Marewski, J. N., Pohl, R. F., & Vitouch, O. (2011b). Recognition-based judgments and decisions: What we have learned (so far). *Judgment and Decision Making*, *6*, 359–380.
- Marewski, J. N., Schooler, L. J., & Gigerenzer, G. (2010). Five principles for studying people’s use of heuristics. *Acta Psychologica Sinica*, *42*, 72–87.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.
- McCloy, R., Beaman, C. P., & Smith, P. T. (2008). The relative success of recognition-based inference in multichoice decisions. *Cognitive Science*, *32*, 1037–1048.
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42–54.
- Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences*, *9*, 11–15.
- Newell, B. R., Collins, P., & Lee, M. D. (2007). Adjusting the spanner: Testing an evidence accumulation model of decision making. In D. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 533–538). Austin, TX: Cognitive Science Society.
- Newell, B. R., & Fernandez, D. (2006). On the binary quality of recognition and the inconsequentiality of further knowledge: Two critical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, *19*, 333–346.
- Newell, B. R., & Lee, M. D. (2010). The right tool for the job? Comparing an evidence accumulation and a naive strategy selection model of decision making. *Journal of Behavioral Decision Making*, DOI: 10.1002/bdm.703.
- Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 923–935.
- Oeusoonthornwattana, O., & Shanks, D. R. (2010). I like what I know: Is recognition a non-compensatory de-

- terminer of consumer choice? *Judgment and Decision Making*, 5, 310–325.
- Oppenheimer, D. M. (2003). Not so fast! (and not so frugal!): Rethinking the recognition heuristic. *Cognition*, 90, B1-B9.
- Pachur, T. (2010). Recognition-based inference: When is less more in the real world? *Psychonomic Bulletin & Review*, 17, 589–598.
- Pachur, T., & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica*, 125, 99–116.
- Pachur, T., Bröder, A., & Marewski, J. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making*, 21, 183–210.
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 983–1002.
- Pachur, T., Mata, R., & Schooler, L. J. (2009). Cognitive aging and the adaptive use of recognition in decision making. *Psychology and Aging*, 24, 901–915.
- Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. (in press). When is the recognition heuristic an adaptive tool? In P. Todd, G. Gigerenzer & The ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world*. New York: Oxford University Press.
- Pleskac, T. J. (2007). A signal detection analysis of the recognition heuristic. *Psychonomic Bulletin & Review*, 14, 379–391.
- Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, 19, 251–271.
- Reimer, T., & Katsikopoulos, K. V. (2004). The use of recognition in group decision-making. *Cognitive Science*, 28, 1009–1029.
- Richter, T., & Späth, P. (2006). Recognition is used as one cue among others in judgment and decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 150–162.
- Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236.
- Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, 112, 610–628.
- Smithson, M. (2010). When less is more in the recognition heuristic. *Judgment and Decision Making*, 5, 230–244.
- Snook, B., & Cullen, R. M. (2006). Recognizing National Hockey League greatness with an ignorance-based heuristic. *Canadian Journal of Experimental Psychology*, 60, 33–43.
- Tomlinson, T., Marewski, J. N., & Dougherty, M. (2011). Four challenges for cognitive research on the recognition heuristic and a call for a research strategy shift. *Judgment and Decision Making*, 6, 89–99.