

Homo heuristicus outnumbered: Comment on Gigerenzer and Brighton (2009)

Benjamin E. Hilbig

University of Mannheim

and Max-Planck Institute for Research on Collective Goods

Tobias Richter

University of Cologne

Please address correspondence to:

Benjamin E. Hilbig

Psychology III, University of Mannheim

Schloss Ehrenhof Ost, 68131 Mannheim, Germany

hilbig@psychologie.uni-mannheim.de

ABSTRACT

Gigerenzer and Brighton (2009) have argued for a ‘homo heuristicus’ view of judgment and decision making, claiming that there is evidence for a majority of individuals using fast and frugal heuristics. In this vein, they criticize previous studies which tested the descriptive adequacy of some of these heuristics. Additionally, they provide a re-analysis of experimental data on the recognition heuristic which allegedly supports Gigerenzer and Brighton’s view of pervasive reliance on heuristics. However, their arguments and re-analyses are both conceptually and methodologically problematic. We provide counterarguments and a re-analysis of the data considered by Gigerenzer and Brighton. Results clearly replicate previous findings which are at odds with the claim that simple heuristics provide a general description of inferences for a majority of decision makers.

Keywords: fast and frugal heuristics, adaptive toolbox, recognition heuristic, formal modeling, multinomial processing tree model

Homo heuristicus outnumbered: Comment on Gigerenzer and Brighton (2009)

In their review of work on the adaptive toolbox of fast and frugal heuristics, Gigerenzer and Brighton (2009) provided a critical discussion of empirical evidence and the methodology that has been used to investigate the assumed non-compensatory nature of these heuristics. One cornerstone of their discussion is a re-analysis of data from an Experiment by Richter and Späth (2006, Experiment 3) on the use of recognition and further knowledge in comparative judgments. In this experiment, in which German students were presented with pairs of names of US-American cities with the task to choose the more populous city, recognition and task-relevant knowledge were varied. In line with the predictions of the recognition heuristic (Goldstein & Gigerenzer, 2002), participants mostly chose recognized cities over unrecognized ones. However, these recognition effects were partly compensated by task-relevant knowledge which conflicts with the claim of non-compensatory reliance on recognition.

Whereas Richter and Späth concluded from their results that recognition information is not generally used in a non-compensatory fashion but integrated with further knowledge (for similar conclusions, see Bröder & Eichler, 2006; Newell & Fernandez, 2006; Pohl, 2006), Gigerenzer and Brighton arrive at a contrary interpretation of the data. They argue that, when analyzed appropriately at the individual level, the data show “that a majority of participants consistently followed the recognition heuristic in the presence of conflicting cues” (p. 134). We believe this interpretation to be conceptually and methodologically flawed. Given the centrality of the recognition heuristic for the adaptive toolbox approach and the attention that the claim of its non-compensatory nature has attracted in the field, we feel that Gigerenzer and Brighton’s claims should not be left undisputed. In this comment, we will focus on a conceptual ambiguity and, ore importantly, a methodological flaw. With respect to the latter, we will then present a re-analysis of the data from Richter and Späth which is based on a

formal measurement model specifically developed to estimate the degree to which decision makers rely on recognition in a non-compensatory fashion.

Conceptual problem: How can use of the recognition heuristic depend on the recognition validity?

The three experiments reported by Richter and Späth (2006) consistently suggested compensatory use of recognition despite the fact that they refer to heterogeneous domains (population of animal species, safety of air carriers, population of American cities). Yet, Gigerenzer and Brighton (2009) dismissed the evidence from two out of the three experiments as being irrelevant for their theory by arguing that in these experiments, “the recognition validity was unknown or low” (p. 133). However, the original theory of the recognition heuristic (Goldstein & Gigerenzer, 2002) does not entail the explicit assumption that the recognition heuristic is (only) used if the recognition validity in a given domain is high. Of course, Goldstein and Gigerenzer (2002) provide the normative observation that the recognition heuristic is *useful* only if this is the case. However, such a normative fact does not necessarily imply the descriptive claim that the recognition heuristic is *applied* if and only if the recognition validity is high.

This ambiguity notwithstanding, it has been shown empirically that decision makers will indeed refrain from relying on the recognition cue when it is invalid (Hilbig, Erdfelder, & Pohl, 2010; Pohl, 2006). At the same time, it is unclear how such an adaptive reliance on recognition is actually achieved by decision makers (an instance of the 'strategy selection problem', cf. Glöckner & Betsch, in press). At least, it appears that more complex processes beyond the simple search, stopping, and decision rules of the recognition heuristic would be necessary. Stated differently, the much-acclaimed simplicity and precision of the recognition heuristic do wane so long as there is no specification of how exactly decision makers are expected to know the recognition validity (of any possible domain) and, thereby, when to rely

on the recognition heuristic. Given these open questions, it seems somewhat harsh to entirely dismiss experiments which entail low recognition validity.

Methodological problem: Choosing the recognized alternative is not equivalent to using the recognition heuristic

In re-analyzing the data from Richter and Späth (2006, Experiment 3), Gigerenzer and Brighton (2009) focused on critical trials in which one alternative was recognized and the other one was not. They presented the number of trials in which the recognized alternative was chosen by each participant (a measure sometimes called *adherence rate*) broken down by whether or not further knowledge argued against the recognized alternative. Based on these data, they concluded that “even in the critical condition [...], the majority of participants consistently followed the recognition heuristic” (p. 134). To the extent that this is meant to imply “a majority of participants consistently *used* the recognition heuristic”, this conclusion represents a logical fallacy (*petitio principii*) because it presupposes the very assumption that is in dispute, viz. that the recognition heuristic is in fact operative whenever participants choose the recognized over the unrecognized alternative.

In general, it is easy to empirically demonstrate high adherence rates to any heuristic which relies on a cue comprising above-chance-level validity (Hilbig, in press). This is due to the confound between the cue in question and other pieces of information pointing in the same direction: In the experiment by Richter and Späth (2006), participants recognizing a city (e.g., Boston) were also likely to know cues that argue *for* the size of this city (e.g., Boston has well-known universities, Boston is located in the relatively more populated North-East of the US etc.), thus potentially resulting in choice of the recognized city. As a consequence, the adherence rate is not a valid indicator of the extent to which people use the recognition heuristic (RH-use) because it severely overestimates use of non-compensatory heuristics in general (Hilbig, 2008b). For this reason, our own re-analysis of the data relies on the estimate

of RH-use provided by the r-model (Hilbig et al., 2010). This model and the results of our re-analysis are described in the remainder of this comment.

A model-based re-analysis of the data from Richter and Späth (2006, Experiment 3)

A measurement model of the recognition heuristic, named r-model, was recently developed by Hilbig et al. (2010) and is depicted in Figure 1. Based on observed categorical data (i.e. choices), this multinomial processing tree model (Erdfelder et al., 2009) provides estimates of one parameter representing non-compensatory reliance on the recognition cue (probability r) as proposed by the recognition heuristic. Additionally, other parameters which stand for the recognition validity a , the knowledge validity b and the probability of valid guesses g are estimated. The basic idea of the r-model – and its main advantage over measures such as adherence rates – is to disentangle the use of recognition and additional knowledge in comparative judgments concerning pairs where one object is recognized and the other one is not (Figure 1, case C). To this end, the knowledge validity b (i.e., the probability of valid knowledge) is also estimated from judgments concerning pairs where both objects are recognized (Figure 1, case A) and, therefore, the recognition heuristic cannot be applied.

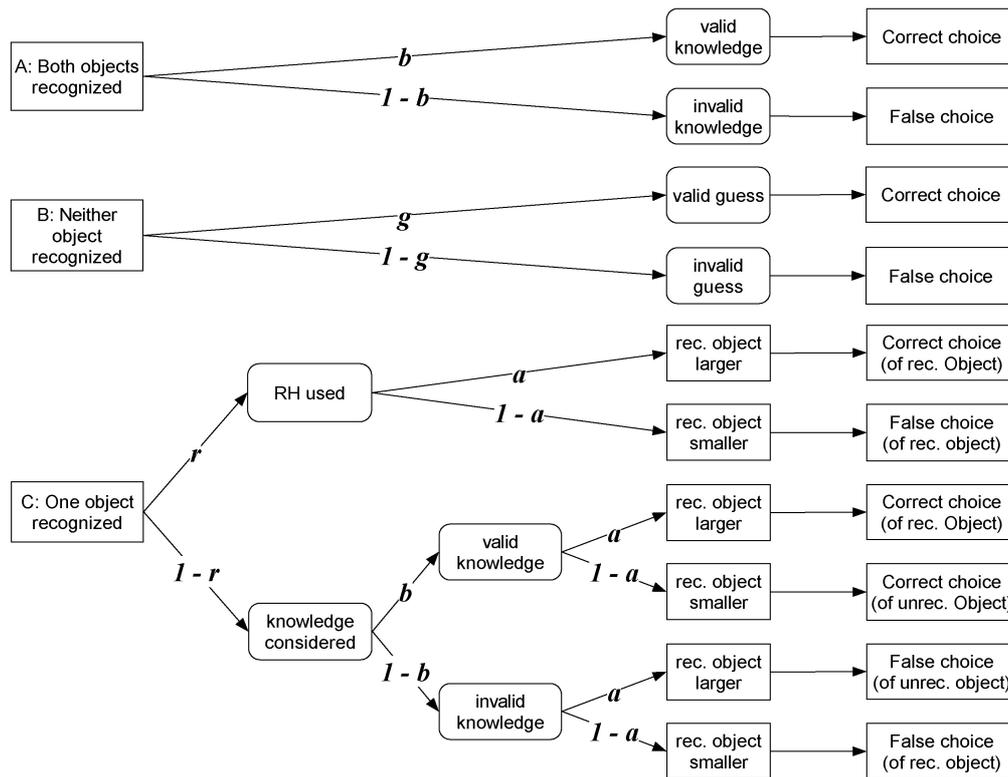


Figure 1. The r-model in the form of processing trees. Three cases are distinguished: (A) both objects are recognized, (B) neither is recognized, or (C) exactly one is recognized. The parameter a represents the recognition validity (probability of the recognized object representing the correct choice), b stands for the knowledge validity (probability of valid knowledge), g is the probability of a correct guess and, most importantly, r denotes the probability of using the recognition heuristic (following the recognition cue while ignoring any knowledge beyond recognition).

The logic of the r-model is simple. Consider a participant who has to make a comparative judgment between two alternatives of which she recognizes only one (Figure 1, case C). In this situation, she can either use the recognition heuristic, which will occur with probability r , or she can consider additional knowledge or information, which will happen with probability $1 - r$. If the participant uses the recognition heuristic and thus chooses the recognized object, her judgment will be correct with probability a , i.e. the recognition validity. If she considers additional knowledge, her judgment will be correct with probability b . In that case, valid knowledge will lead to a correct choice which can, in fact, either mean choosing the recognized or the unrecognized of the two objects – depending on which

represents a correct judgment in the current pair. Within the r -model, the recognition heuristic can be implemented as a submodel by fixing the r parameter to 1. The r -model has been shown to fit empirical data well and the psychological meaning of the central model parameter r has been validated experimentally (Hilbig et al., 2010). Moreover, simulations revealed that the r -model provides the best estimate of RH-use currently available (Hilbig, 2010). Consequently, we used it for a re-analysis of the data from Richter and Späth (2006, Experiment 3) both on the aggregate as well as the individual level.

Method

In Richter and Späth's data, we determined for each pair of cities presented to participants which of the two cities a participant had reported to recognize. Each pair could thus be sorted into one of the three trees in the r -model (cases A, B, or C, respectively). Next, it was determined for each pair which of the two cities represented the factually correct option with respect to the judgment criterion, that is, city population. Thereby, each choice could be classified as correct or false. Finally, in cases in which only one city was reported to be recognized, we determined whether the recognized (or the unrecognized) of the two cities had been chosen, that is, judged to be more populous. With these three steps, every choice in Richter and Späth's data was sorted into one of the eight possible outcome categories shown in the terminal branches of Figure 1.

As is most typically the case in multinomial modeling, parameter estimates were sought by minimizing the asymptotically χ^2 -distributed log-likelihood ratio statistic G^2 through the EM-algorithm (Hu & Batchelder, 1994). In a nutshell, this maximum-likelihood procedure searches through the parameter-space to determine the set of parameters which minimizes the distance between observed and expected category frequencies (in the current case, the eight choice categories described above). Parameter estimates and model fit statistics for the r -model were obtained using the multiTree software tool (Moshagen, 2010). Model fits

were tested by means of the goodness-of-fit statistic G^2 and differences between nested models with the according χ^2 -difference test for changes in model fit (ΔG^2). Non-nested models were compared using the Bayesian Information Criterion (BIC; e.g. Myung, 2000) from which the Bayesian posterior probability of model superiority (given the data) can be computed, assuming equal priors (Wagenmakers, 2007).

Results and Discussion

Aggregate analyses

On the aggregate level (across all choices and participants) the r-model accounted for the data well, as shown by a satisfactory fit of $G^2(1) = 1.6, p = .20$. The obtained parameter estimates were $a = .88 (SE = .01)$, $b = .59 (SE = .01)$, $g = .49 (SE = .02)$ and, most importantly, $r = .80 (SE = .01)$. As such, the estimated probability of true RH-use was substantial, though significantly smaller than implied by the adherence rate ($M = .91$), $\Delta G^2(1) = 109.5, p < .001$. Once again, this finding confirms that RH-use is overestimated by adherence rates.

As the previous findings imply, a strict and deterministic version of the recognition heuristic (fixing $r = 1$) also failed to account for the data and produced severe misfit ($p < .001$). However, such a deterministic understanding of the recognition heuristic may be seen as unfair, since strategy execution errors must be expected (e.g. Bröder & Schiffer, 2003). As a consequence, we next implemented the recognition heuristic in a probabilistic rather than a deterministic way (cf. Hilbig et al., 2010): First, we added an error parameter f to each terminal branch of the original r-model, thus implementing a naïve error theory (Rieskamp, 2008). This extended r-model was then compared to a submodel with r fixed at 1 which represents a probabilistic version of the recognition heuristic. Comparing these models revealed that the probabilistic recognition heuristic submodel needed an average error of $f = .09 (SE = .01)$ to account for the data. Nevertheless, it still fit the data worse than the extended

r-model, $\Delta G^2(1) = 10.1$, $p = .001$. As such, even a probabilistic version of the recognition heuristic could not account for the data as well as a model implying that the recognition cue is only sometimes considered in isolation ($r < 1$).

Finally, we compared the original r-model (without any error parameter) to the probabilistic recognition heuristic submodel (as before, fixing $r = 1$ and adding an error parameter f). Since these are non-nested models, we based the model comparison on the BIC which was 7119 for the original r-model and 7129 for the probabilistic recognition heuristic submodel. As such, the r-model was superior, while both comprise the exact same number of free parameters; the Bayesian posterior probability (given the data, assuming equal priors) of the r-model as compared to the probabilistic implementation of the recognition heuristic was .99 which can be understood as very strong evidence against the latter (Wagenmakers, 2007).

In sum, non-compensatory reliance on recognition did occur in a substantial proportion of cases which is plausible given the extremely high recognition validity. However, the model-based aggregate analyses revealed that the data could not be adequately explained by the recognition heuristic, not even when implemented probabilistically. A model in which the recognition cue is only sometimes considered in isolation generally fit the data better. This model is in line with compensatory theories which propose that the recognition cue is integrated with others (if available). The findings thus mirror previous investigations with other data sets (Hilbig et al., 2010) and confirm Richter and Späth's (2006) original conclusions, that even though recognition is indubitably a very prominent cue it is 'used as one cue among others'.

Individual analyses

In accordance with the arguments put forward by Gigerenzer and Brighton (2009) we next analyzed the choice data of each individual separately, again using the r-model. The results are displayed in Figure 2 in which the grey bar indicates the corresponding individual

estimate of r as compared to the individual adherence rate (white bar) for each participant. As can be seen, RH-use was less likely than implied by the adherence rate for practically every participant. Stated differently, only a small number of participants consistently relied on the recognition cue in a non-compensatory manner. Most participants, by contrast, refrained from doing so in a non-trivial proportion of cases which was clearly lower than implied by their individual adherence rates – the measure on which Gigerenzer and Brighton (2009) based their conclusions.

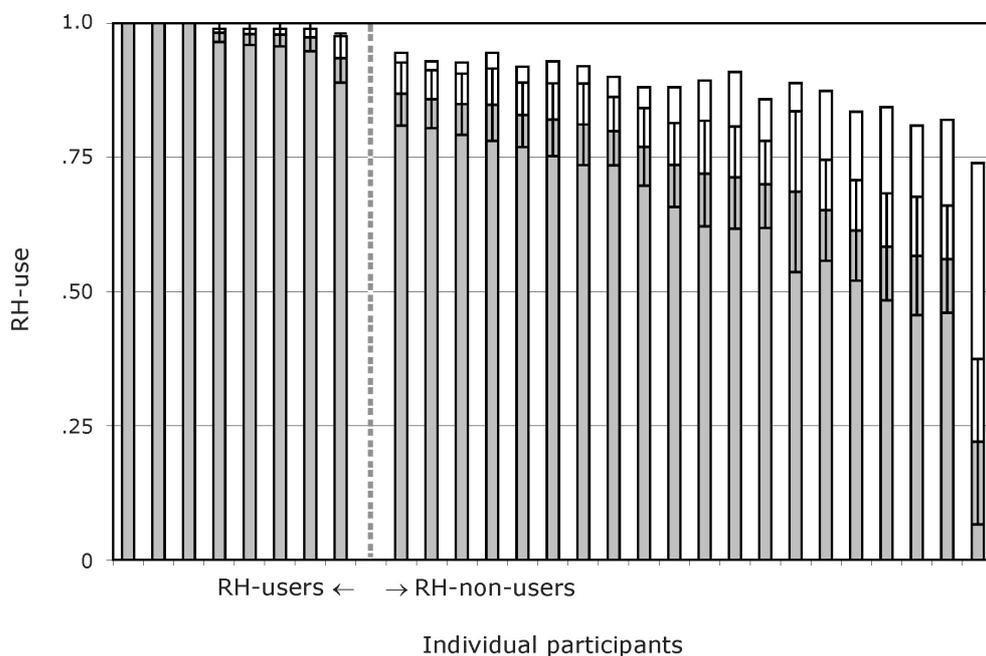


Figure 2 Individual probability of RH-use as estimated by the r -parameter (grey bar including one standard error of the parameter estimate, cf. Moshagen, 2010) and by the individual adherence rate to the predictions of the recognition heuristic (white bar).

To test, on the individual level, how many participants might be classified as users of the recognition heuristic, we first used the procedure described in Hilbig et al. (2010): Taking the average parameter estimates as the alternative hypothesis (H_1) we performed a power analysis (Erdfelder, Faul, & Buchner, 2005) to determine the value of r_0 which implied a power of $1 - \beta = .95$ for testing the original r -model against the recognition heuristic submodel

with r_0 fixed accordingly. The resulting parameter value of r_0 was .96. So, for each participant, we compared the fit of the original r -model (with no constraint on r) to a submodel fixing $r = .96$ which represents the recognition heuristic. Comparing the models showed that for 20 out of the 28 participants (71%, see Figure 1), the submodel of the recognition heuristic fit significantly ($p < .05$) worse than the r -model. For these participants, the probability of using recognition was significantly smaller than the critical value of .96. Stated differently, this majority of decision makers too rarely used the recognition heuristic to be classified as consistent RH-users.

However, one may once more argue that $r = .96$ (without any strategy execution error) is too strict an implementation of the recognition heuristic. Therefore, mirroring the aggregate analyses reported above, we compared the original r -model (no constraint on r , no error parameter) to the probabilistic recognition heuristic submodel (fixing $r = 1$ and adding an error parameter f) on the individual level. Again, these are non-nested models which were consequently compared using the BIC. We found that for 9 participants (32%) the probabilistic recognition heuristic was the superior model (yielding the smaller BIC value), for 7 (25%) neither model performed better, and for 12 (43%) the r -model was to be preferred. So, even when implementing a probabilistic version of the recognition heuristic and comparing models at the individual level, more participants were classified as RH-non-users than as RH-users.

Finally, additional evidence was obtained from a model-free analysis using the individual discrimination index (DI; Hilbig & Pohl, 2008) which is defined as the difference in adherence rates depending on whether using recognition implies a correct versus a false inference. Any participant reliably discriminating such cases cannot have relied on recognition alone. Analyses revealed that 11 participants (39%) had a DI score within the 95%-confidence interval of zero, thus qualifying as *potential* users of the recognition heuristic. Vice versa, the remaining 17 participants consistently discriminated whether

recognition led to a correct inference or not which is incompatible with the assumption of one-reason decision making as implied by the recognition heuristic (Hilbig, Pohl, & Bröder, 2009).

Conclusions

In our comment, we focused on a conceptual and a methodological drawback inherent in Gigerenzer and Brighton's (2009) critique of Richter and Späth (2006). We argued that the recognition heuristic cannot simply be taken to depend on the recognition validity without further specification, and that adherence to recognition is not a valid measure of RH-use. In order to obtain a valid and unbiased estimate of RH-use, we applied the r-model (Hilbig et al., 2010) for a re-analysis of Richter and Späth's (Experiment 3) data. Both aggregate- and individual-level results showed that the recognition heuristic cannot adequately account for choice data – which held for the majority of participants. These findings are in line with previous experiments all of which cast doubts on the recognition heuristic and other heuristics as general accounts of judgment and decision making (for an overview see Hilbig, in press). This is especially note-worthy given that Gigerenzer and Brighton (2009) consider these data “the perfect test for the heuristic” (p. 133). Indeed, with a recognition validity of .88 in the current data set, it is hard to imagine how any further cues should override the recognition cue particularly often. Given the large recognition validity and (by comparison) low knowledge validity of .59, most alternative models (e.g. Glöckner & Betsch, 2008a; Newell & Lee, in press) must necessarily predict choices which frequently resemble RH-use (Glöckner, 2009). Even so, the assumption of consistent non-compensatory reliance on the recognition cue was rejected for a majority of participants.

Of course, the results also show that RH-use does occur in a substantial proportion of cases. Likewise, there are individuals who seem more prone to apply this strategy (cf. Hilbig, 2008a). As a consequence, further research is needed to uncover its situational and individual

determinants. Such a research agenda has been fruitfully pursued in the past for other heuristics (Bröder, in press; Bröder & Newell, 2008).

To conclude, advertising pervasive use of fast and frugal heuristics is simply not warranted given the empirical data (Hilbig, in press). Re-analyses of selected experimental conditions from single studies using biased measures are unlikely to change this fact. As alternatives to the adaptive toolbox approach, several process models have been suggested (for an overview see Glöckner & Witteman, 2010) and successfully tested (e.g. Glöckner & Betsch, 2008b; Hilbig & Pohl, 2009; Newell & Lee, in press). Thus, we are wary of accepting a ‘homo heuristicus’ view of human decision making, given that fast and frugal heuristics are only used consistently by a minority of decision makers.

REFERENCES

- Bröder, A. (in press). The quest for Take The Best: Insights and outlooks from experimental research. In P. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world*. New York: Oxford University Press.
- Bröder, A., & Eichler, A. (2006). The use of recognition information and additional cues in inferences from memory. *Acta Psychologica, 121*, 275-284.
- 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). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making, 16*, 193-213.
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie - Journal of Psychology, 217*, 108-124.
- Erdfelder, E., Faul, F., & Buchner, A. (2005). Power analysis for categorical methods. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3, pp. 1565-1570). Chichester, UK: Wiley.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*, 107-143.
- Glöckner, A. (2009). Investigating intuitive and deliberate processes statistically: The multiple-measure maximum likelihood strategy classification method. *Judgment and Decision Making, 4*, 186-199.
- 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, & Cognition*, *34*, 1055-1075.
- Glöckner, A., & Betsch, T. (in press). Accounting for critical evidence while being precise and avoiding the strategy selection problem in a parallel constraint satisfaction approach – a reply to Marewski (in press). *Journal of Behavioral Decision Making*.
- Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorization of processes underlying intuitive judgment and decision making. *Thinking & Reasoning*, *16*, 1-25.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75-90.
- Hilbig, B. E. (2008a). Individual differences in fast-and-frugal decision making: Neuroticism and the recognition heuristic. *Journal of Research in Personality*, *42*, 1641-1645.
- Hilbig, B. E. (2008b). One-reason decision making in risky choice? A closer look at the priority heuristic. *Judgment and Decision Making*, *3*, 457-462.
- Hilbig, B. E. (2010). *Precise models deserve precise measures: a methodological dissection*. Manuscript submitted for publication.
- Hilbig, B. E. (in press). Reconsidering ‘evidence’ for fast and frugal heuristics. *Psychonomic Bulletin & Review*.
- 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, & Cognition*, *36*, 123-134.
- 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.
- Hu, X., & Batchelder, W. H. (1994). The statistical analysis of engineering processing tree models with the EM algorithm. *Psychometrika*, *59*, 21-47.
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42-54.
- Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*, 190-204.
- 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. (in press). The right tool for the job? Comparing an evidence accumulation and a naïve strategy selection model of decision making. *Journal of Behavioral Decision Making*.
- Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, *19*, 251-271.
- 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. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1446-1465.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779-804.