TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 3 (2011) 197–205

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2010.01124.x

Towards Competitive Instead of Biased Testing of Heuristics: A Reply to Hilbig and Richter (2011)

Henry Brighton, Gerd Gigerenzer

Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development

Received 8 September 2010; received in revised form 4 November 2010; accepted 4 November 2010

Abstract

Our programmatic article on *Homo heuristicus* (Gigerenzer & Brighton, 2009) included a methodological section specifying three minimum criteria for testing heuristics: competitive tests, individual-level tests, and tests of adaptive selection of heuristics. Using Richter and Späth's (2006) study on the recognition heuristic, we illustrated how violations of these criteria can lead to unsupported conclusions. In their comment, Hilbig and Richter conduct a reanalysis, but again without competitive testing. They neither test nor specify the compensatory model of inference they argue for. Instead, they test whether participants use the recognition heuristic in an unrealistic 100% (or 96%) of cases, report that only some people exhibit this level of consistency, and conclude that most people would follow a compensatory strategy. We know of no model of judgment that predicts 96% correctly. The curious methodological practice of adopting an unrealistic measure of success to argue against a competing model, and to interpret such a finding as a triumph for a preferred but unspecified model, can only hinder progress. Marewski, Gaissmaier, Schooler, Goldstein, and Gigerenzer (2010), in contrast, specified five compensatory models, compared them with the recognition heuristic, and found that the recognition heuristic predicted inferences most accurately.

Keywords: Simple heuristics; Recognition heuristic; *Homo heuristicus*; Biased testing

1. Introduction

Cognition rests on an ability to make accurate inferences from limited observations of an uncertain and potentially changing environment. Developing theories capable of explaining how the cognitive system functions so effectively despite this uncertainty is a key step

Correspondence should be sent to Henry Brighton, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: hbrighton@mpib-berlin.mpg.de

toward understanding cognition. The abilities of machines, for example, pale in comparison. These issues drive our research, and the notion of *Homo heuristicus* characterizes a particular relationship between cognition and the structure of the environment, one that hypothesizes how an organism can make accurate inferences about an uncertain world (Gigerenzer & Brighton, 2009). Rather than attempting to minimize, maximize, or optimize during the process of problem solving, *Homo heuristicus* relies on heuristic, resource-frugal, and robust solutions that ignore information. This does not mean that heuristics are functionally inferior to processes that integrate all information or optimize. Optimization is not always possible or desirable in the complex and uncertain environments in which we find ourselves. In fact, a mind that relies on simple heuristics can make both faster and more accurate inferences than one that relies on, for example, multiple regression (Czerlinski, Gigerenzer, & Goldstein, 1999) or neural networks models (Brighton, 2006; Chater, Oaksford, Nakisa, & Redington, 2003). Our article examined less-is-more effects and used the statistical problem of the bias-variance dilemma to further understand how, when, and why heuristics make accurate inferences.

More specifically, the study of *Homo heuristicus* proceeds by (a) proposing heuristics, expressed as precise and testable computational models, their building blocks, and the core cognitive capacities they exploit; (b) analyzing the functional–ecological implications of these heuristics, which means understanding why and when they work; (c) examining how heuristics are selected from what we refer to as an *adaptive toolbox*, a metaphor used to conceptualize the stock of heuristics available to the organism. Consequently, empirical tests of heuristic use should be guided by knowledge of their functional–ecological implications, because the hypothesis is that people will attempt to select specific heuristic when it is adaptive to do so. By examining questions like these, we aim to lay firm foundations for understanding the broader issue of strategy selection.

The problem of strategy selection is not specific to the study of heuristics. It should be a concern for anyone who accepts that cognition, and therefore decision making, relies on more than one form of processing (e.g., Einhorn, 1970, 1971; Ginossar & Trope, 1987; Payne, 1976; Payne, Bettman, & Johnson, 1988; Payne, Bettman, & Johnson, 1993; Rapoport & Wallsten, 1972; Rieskamp & Otto, 2006; Svenson, 1979). Hilbig and Richter remark that the simplicity of heuristics wanes once the task of strategy selection has been taken into account. This criticism assumes that the problem of strategy selection demands complex processes, although no supporting evidence for this assumption was given. It also implies that the strategy selection problem represents the Achilles heel of research into heuristics, while other approaches can, somehow, safely disregard the problem. This strikes us as short-sighted. Hillbig and Richter's viewpoint requires a commitment to believing that a single strategy, or a single pattern of information processing, underlies the problem of inductive inference. This line of reasoning places the onus firmly on those adopting such a view to explain how a single process could adequately respond to the diversity of statistical patterns found in the environment (e.g., Geman, Bienenstock, & Doursat, 1992; Schaffer, 1993). Moreover, for those who assume that the mind has only one tool in its adaptive toolbox, such as a weighted-linear strategy or a neural network, the strategy selection problem translates into the question of how the mind selects a new and adequate parameter set for every new class of problem.

2. Three methodological principles for testing strategies

Hilbig and Richter are largely mute on these theoretical issues. Instead, they respond to our critique of an experiment by Richter and Späth (2006) on the recognition heuristic. In Section 5, on Methodology and Empirical Evidence, we put forward three methodological principles for testing heuristics (Gigerenzer & Brighton, 2009, p. 132). We then used the Richter and Späth study to illustrate how violating these principles can lead to unwarranted conclusions. The three principles are:

(i) competitive tests,
(ii) individual-level tests, and
(iii) tests of adaptive selection of heuristics.

Richter and Späth's study violated all three principles. First, the authors tested only the recognition heuristic but concluded that a compensatory strategy they had neither tested nor specified would explain participants' inferences more accurately. Second, this conclusion was based on means only; no individual data were analyzed. Note that in the presence of systematic individual differences, one should not draw conclusions about individual processes from group means. In the extreme, the mean will not represent a single individual. Third, the authors tested if subjects used the recognition heuristic without first establishing if the statistical properties of the task—such as the presence of substantial recognition validity—made it functional to do so. In Experiment 1, the recognition validity was not reported and likely at chance. In Experiment 2, an adjusted small correlation was reported instead of the recognition validity. Only in Experiment 3 was the recognition validity substantial. The adaptive selection of heuristics would predict that accordance rates are high in Experiment 3, but low or at chance level in the others. In contrast, Richter and Späth appear to have missed the point of the adaptive selection of heuristics, and the study of ecological rationality (Gigerenzer et al., 1999), which we spent nine pages discussing in our article (Gigerenzer & Brighton, 2009, p. 116–125). Richter and Späth (2006, p. 160) went as far to suggest that the recognition heuristic is ''universally applied'' or that people ''rely on recognition blindly.'' Hilbig and Richter (p. 4) perpetuate this misunderstanding of heuristics as general-purpose rules, and even attribute it to Goldstein and Gigerenzer (2002), despite these authors explicitly stating that the recognition heuristic is not a general purpose strategy. The ecological rationality of the recognition heuristic is defined by two characteristics: Some objects must be unrecognized *and* the recognition validity must be substantial (pp. 76–78, 87).

We fully accept that the details of how the cognitive system shifts strategies adaptively in response to recognition validity and other factors has not been fully set out. We do not accept that one should refrain from using knowledge of the functional–ecological implications of a model to inform the process of experimental design and analysis. While Hilbig and Richter consider it ''somewhat harsh'' (p. 5) to doubt the insights of a study for which the functional–ecological match between the task and the recognition heuristic is weak or unknown, we consider it absolutely central to conducting solid empirical work.

2.1. Richter and Späth's (2006) study: No competitive testing, no individual analyses

To avoid any further misunderstanding, let us first define the adaptive use of the recognition heuristic. Relying on the heuristic is ecologically rational in an environment R where the recognition of objects $a, b \in R$ is strong and positively correlated with their criterion values. For two objects, the heuristic is:

> If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion.

This heuristic is noncompensatory in the sense that the recall of further cues about the recognized object cannot compensate for (i.e., overturn) recognition information. In the original work (Gigerenzer & Goldstein, 1996, pp. 651–652; Goldstein & Gigerenzer, 2002, pp. 76–78), the recognition heuristic was assumed to model human inferences when three conditions hold:

(i) the recognition validity is substantial,
(ii) inferences are made from memory, rather than from tables of information, and
(iii) recognition stems from natural environments, as opposed to artificial manipulation.

In Experiment 3 of Richter and Späth (2006), these conditions were fulfilled. The authors then asked whether the recognition heuristic predicts people's inferences in the presence of a strong, contradicting cue. German participants were taught whether certain recognized American cities have international airports or not. The airport cue was chosen as being the most valid (mean subjective validity = .82) among six cues tested in a pilot study. Moreover, the biserial rank correlation between population rank and airport was *larger* than that between population rank and recognition, .71 versus −.56. There were three memory states for recognized cities: positive cue (with international airport), no cue (unknown), and negative cue (no international airport). Richter and Späth reported that in these three states, 98%, 95%, and 82% of all inferences were in accordance with the recognition heuristic, respectively. Their conclusion, though, was remarkable: ''no evidence was found in favor of a noncompensatory use of recognition'' (p. 159).

We presented an analysis of Richter and Späth's data at the individual level, which showed that even in the presence of a strong contradicting cue, the *majority of participants* (17 out of 28) chose the recognized objects all the time (32 out of 32 judgments) or nearly all the time (31 out of 32), while the others appeared to guess or follow some other strategy (Gigerenzer & Brighton, 2009, figure 7). This pattern showed a degree of intraindividual consistency rarely obtained in judgment and decision-making research. Richter and Späth (2006, p. 159), in contrast, had concluded that there would be ''clear evidence'' for compensatory strategies they favor, without having formulated and tested such a model.

3. Hilbig and Richter's reanalysis: Still no competitive testing

Before we comment on Hilbig and Richter's reanalysis, we would like to correct three errors repeated throughout their article.

1. *Not all heuristics are noncompensatory processes.* Hilbig and Richter begin their comment by stating, ''Gigerenzer and Brighton (2009) provided a critical discussion of empirical evidence and the methodology that has been used to investigate the assumed noncompensatory nature of these heuristics'' (p. 3). The claim that all heuristics are noncompensatory processes is incorrect. In fact, the first example of a heuristic we discussed at length was tallying, a compensatory heuristic. This oversight allows the authors to make a second erroneous claim.
2. *If a person does not rely on the recognition heuristic, it does not follow that he or she relies on a nonheuristic compensatory strategy.* Hilbig and Richter interpret the finding that not all participants follow the recognition heuristic as evidence that they follow a ''nonheuristic'' compensatory strategy. This conclusion is invalid, for two reasons. First, as mentioned before, participants may rely on a compensatory heuristic such as tallying. Second, participants may rely on a different noncompensatory heuristic, such as a one-reason heuristic that only considers information about international airports. The point is, an argument for an alternative explanation that has not been formalized as a model and tested will necessarily be an argument based on speculation.
3. *To generalize from one heuristic to all heuristics is logically unfounded.* Hilbig and Richter claim in their abstract and conclusions that ''fast and frugal heuristics are only used consistently by a minority of decision makers'' (p. 14). It should be clear, though, that such statements about heuristics in general cannot be justified after analyzing one heuristic, as is the case in their comment.

In their reanalysis of Richter and Späth's data, Hilbig and Richter use an individual-level analysis, but no competitive testing. Moreover, lack of competitive testing is combined with biased testing of the recognition heuristic, which is our next point.

3.1. The 100% (96%) threshold

Hilbig and Richter reanalyzed the data of Richter and Späth (2006) using the *r*-model. This model attempts to estimate the proportion of people that ''use'' the recognition heuristic, an estimate which they propose as being more accurate than the adherence rate. The adherence rate, the *r*-model, or any other criterion one chooses to use, can only ever provide an uncertain indicator of the ''use'' of a cognitive model. Given this, we are nonplussed by Hilbig and Richter's accusation that we make a logical error when using adherence rate to test the predictive accuracy of the recognition heuristic. The issue is not, and never will be, one of logic. Science, fortunately, is clear on how to resolve the matter: Competing explanations should be judged on their ability to explain the data. The problem is, the *r*-model is a model of behavior and in no way specifies, beyond the recognition heuristic, how information is processed when making decisions. It fails to offer a valid competing explanation in what should be competition played on a level playing field between process models, models that attempt to describe the data-generating machinery.

Rather than delve into the specifics of their analysis and quibble over the ability of a behavioral model to offer solid insights into cognitive processing, we will instead focus on the criterion used by Hilbig and Richter to classify a person as a user of the recognition heuristic. We use the term ''biased'' testing if someone evaluates two or more process models in competitive testing but uses different evaluation criteria. In the absence of competitive testing, biased testing is the practice of assessing the model one disfavors against an unrealistic standard. The first test Hilbig and Richter conduct with the *r*-model uses a 100% threshold ($r = 1$); that is, they test the hypothesis that *every* person *always* relies on the recognition heuristic. Then they relax 100% to 96%, and classify only a minority (9 out of 28 participants in Richter and Späth's Experiment 3) as ''users'' of the recognition heuristic. (These are likely to be the same nine participants who followed the predictions of the recognition heuristic in 32 out of 32 times, according to our reanalysis in Gigerenzer & Brighton, 2009, figure 7, lower panel.)

It should be evident that by choosing any number, say, 100%, 90%, or 75% of subject's responses as a threshold, or augmenting the *r*-model with an error-theory, one can obtain more or less favorable results for the recognition heuristic. The particular threshold values Hilbig and Richter chose are not met by any model of cognitive processes we are aware of in the entire field of judgment and decision making. Prospect theory, a Nobel Prize-model, for instance, typically predicts 75%–80% of judgments in two–alternative choice tasks. Other models do not much better, and often worse (Brandstätter, Gigerenzer, & Hertwig, 2006). In other work, where Hilbig, Scholl, and Pohl (2010) estimate the *r*-value rather than fix it at an unrealistic level, they conclude that the majority of judgments, ranging from 63% to 77%, resulted from ''use'' of the recognition heuristic.

To summarize, the *r*-model analysis of Hilbig and Richter does not specify an alternative process model or test it competitively with the recognition heuristic. They attempt to estimate the ''true'' proportion of recognition heuristic users. The resulting estimated proportion depends on an arbitrary threshold, which the authors set unrealistically high as 100% or 96%. Without carrying out the additional work of specifying a competing model, it is impossible to know how another process model would fare when measured against this same criterion. Competitive model testing renders arbitrary decision thresholds such as these irrelevant and provides clear answers to questions concerning the relative ability of models to explain behavior.

4. How to resolve the debate: Competitive testing

Hilbig and Richter clearly harbor intuitions about a superior explanation for human decision making. By abstaining from the challenge of formulating and putting their intuitions to the test, they are free to enjoy the luxury of speculation. We welcome competing proposals and see them as essential to progress, but unless these intuitions are formalized to an extent that they can be compared with existing models and judged on the same footing, then these intuitions will remain intuitions. They should not be mistaken for a serious competing explanation, and they should certainly not be used as a means to arrive at evidence against existing models that *have* been formally specified and can undergo empirical testing.

How can this debate be resolved? The answer is simple. Specify an alternative model, and then assess the relative ability of competing models to explain the observations. This is why we have stressed the competitive testing of process models in the title of this response, and in the abstract of the original article. Such formal models exist. Already in the first article on the recognition heuristic, Gigerenzer and Goldstein (1996) formulated and analyzed the predictive accuracy of several compensatory models, including variations on tallying that use recognition information as ''just another cue,'' as Hilbig and Richter argue.

Of particular relevant to this debate is the first experimental study that competitively tests various cognitive process models that assume compensatory processing of recognition information. Marewski et al. (2010) formulated five alternative models that integrate recognition with further cues for the recognized object. All five competing models can be thought of as weighted linear additive models with two classes of predictors, recognition and cues, or recognition and retrieval time, respectively. These models have free parameters that allow them to mimic the recognition heuristic *and* predict the opposite pattern, depending on the parameter values. That is, they included the recognition heuristic as a special case. None of the five compensatory models could predict judgments with greater accuracy than the recognition heuristic, which performed best overall. The study shows that although the recognition heuristic cannot predict with 100% accuracy, particularly in the presence of contradicting cues, this finding by itself does not imply that compensatory models predict with greater accuracy.

5. Conclusion

We introduced the notion of *Homo heuristicus* to characterize how the cognitive system might make accurate inferences in uncertain, complex, and potentially changing environments. Our hypothesis is that by ignoring information, such as cues and dependencies between cues (a form of bias), organisms can simultaneously achieve robust, functional, and tractable responses to environmental uncertainty. Heuristics are process models used to formalize and test this hypothesis. In contexts where all events, actions, and probabilities are known, the process of optimization is all well and good, so long as it is computationally tractable to implement. In contexts where optimization is not tractable, or when full knowledge of the problem is unavailable, heuristics can offer superior responses to uncertainty.

In their commentary, Hilbig and Richter have little to say on these functional–ecological issues, issues that play a critical role in constraining cognitive theorizing. Instead, they focus on a specific issue, the proportion of participants who relied on the recognition heuristic in a study by Richter and Späth (2006). Whereas Richter and Späth had concluded that ''no evidence was found'' (p. 159) in favor of the recognition heuristic, the *r*-model reanalysis by Hilbig and Richter now classifies 29% of participants as users of the recognition heuristic. This classification is, however, based on an arbitrary and absolutely unrealistic 96% threshold for classifying participants as users of the recognition heuristic. Based on this threshold, they argue, inaccurately, that some compensatory and therefore ''nonheuristic'' process provides a superior explanation of human behavior, but they fail to specify this alternative and test it.

The basic practice of competitive model testing renders this suspect methodological practice unnecessary and allows competing explanations to vie on a level playing field. To peddle putative evidence against one process model, based on an arbitrary decision criterion, as evidence for an unspecified alternative strikes us as a particularly weak form of argument. To then extend this pattern of reasoning in an attempt to dismiss an entire class of models, such as heuristics in general, strikes us as wholly unconvincing.

References

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without tradeoffs. *Psychological Review*, *113*, 409–432.

Brighton, H. (2006). Robust inference with simple cognitive models. In C. Lebiere & R. Wray (Eds.), *AAAI Spring Symposium: Cognitive Science Principles Meet AI-Hard Problems* (pp. 17–22). Park, CA: American Association for Artificial Intelligence.

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, *90*, 63–86.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, *73*, 221–230.

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

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, *4*, 1–58.

Gigerenzer, G., & Brighton, H. (2009). *Homo heuristicus*: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 155–175.

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

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

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

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.

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.

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 multi-alternative inference. *Psychonomic Bulletin & Review*, *17*, 287–309.

Payne, J. W. (1976). Task complexity and contingent processing in decision processing: An information search and protocol analysis. *Organizational Behavior & Human Decision Processes*, *16*, 366–387.

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

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

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

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.

Schaffer, C. (1993). Overfitting avoidance as bias. *Machine Learning*, *10*, 153–178.

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