

Finding Foundations for Bounded and Adaptive Rationality

Ralph Hertwig¹ · Arthur Paul Pedersen¹

Received: 3 July 2015 / Accepted: 3 July 2015 / Published online: 18 August 2015
© Springer Science+Business Media Dordrecht 2015

In November 2013, we held an interdisciplinary workshop at the Max Planck Institute for Human Development in Berlin entitled “Finding Foundations for Bounded and Adaptive Rationality.” The invited speakers and discussants included psychologists and cognitive, computer, and decision scientists, as well as philosophers; the late Patrick Suppes gave a video presentation from his office at Stanford University. Each presentation had two discussants, one from philosophy and one from the sciences. The discourse that ensued among the workshop’s participants was intensive and constructive, resulting in the eight articles that comprise this special issue of *Minds and Machines*.

In organizing the workshop, we pursued two interrelated goals. The first was to facilitate critical discussion about old and new problems in the study of rationality, particularly those raised and addressed by the *simple-heuristics* program, a research paradigm that, since the mid-1990s, has pursued a novel vision of bounded rationality. The second goal was to transcend the conventional division of labor between behavioral decision scientists and philosophers. Over many decades, the two sets of researchers appear to have agreed on a labor contract. Philosophers are to explicate the nature of rationality and articulate its normative standards. Taking these normative standards lock, stock, and barrel, behavioral decision scientists are then to empirically investigate people’s behavior to ascertain the extent to which

This special issue is dedicated to the memory of Patrick Colonel Suppes (1922–2014).

-
- ✉ Ralph Hertwig
hertwig@mpib-berlin.mpg.de
 - ✉ Arthur Paul Pedersen
pedersen@mpib-berlin.mpg.de

¹ Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

their judgments and decisions conform to or deviate from these standards. Should aberrations occur, psychologists are to diagnose and seek to explain them. The novel vision of bounded rationality pursued by the simple-heuristics program since the mid-1990s has challenged this division, suggesting that it is not conformity to classical canons of rationality that enables sound judgment and good decisions, but the fit of cognitive systems to environmental structures that do.

Bounded Rationality: A Map of Systematic Biases?

Finding choice anomalies has been the *modus operandi* of today's most influential research paradigm in psychology on human judgment and decision making. Pioneered by Daniel Kahneman and Amos Tversky (e.g., Tversky and Kahneman 1974; Kahneman et al. 1982; Kahneman 2011), this research paradigm, called the *heuristics-and-biases* program, has adopted an effective research strategy based on a simple three-step protocol. First, take some principle from logic, probability theory, statistics, or decision theory that many accept to be a normative standard of rationality. Second, determine the extent to which people's judgments and decisions conform to the principle in question. Third, explain deviations from the standard either in terms of heuristics—that is, simple cognitive strategies—or in terms of a “minimal set of modifications” to some theory widely believed to enjoy a normative status, such as expected utility theory (Kahneman 2000, p. x).

Using this protocol, the heuristics-and-biases program has assembled an extensive catalog of putative deviations from supposed canons of rationality (see Conlisk 1996; Krueger and Funder 2004), branding them “biases,” “fallacies,” or “cognitive illusions.” A common appraisal of such deviations has been that the fault resides in people's cognitive system and, in particular, the Janus-faced assortment of simple heuristics it uses (e.g., representativeness, availability, and anchoring-and-adjustment). Although these heuristics are acknowledged to be vital cognitive tools, assisting people with limited cognitive resources to navigate a complex world under conditions of uncertainty, they are also seen to be a liability, engendering systematic deviations from canons of rationality. In Kahneman's (2003) words: “Our research attempted to obtain a map of bounded rationality, by exploring the systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models” (p. 1449). This viewpoint posits a gap between canons of rationality and human performance, and it confines psychology's role to the clinical diagnosis and explanation of those instances in which human judgment and decision making go astray.

In equating systematic biases with bounded rationality, Kahneman (2003) suggested an intellectual kinship between his research program's endeavors to discover cognitive illusions (and the heuristics prone to producing them) and Simon (1956, 1978) vision of bounded rationality. Yet the heuristics-and-biases program, which seeks to clarify the distinction between the *is* and the *ought*, stands in marked opposition to Simon's vision of bounded rationality, which seeks to discern how the *is* sheds light on the *ought*. Guided by Simon's vision, the simple-heuristics program (e.g., Gigerenzer et al. 2011) has substantially enriched psychological research on

human judgment and decision making over the past two decades. What is this alternative interpretation of bounded rationality?

Bounded Rationality: A Toolbox of Ecologically Rational Heuristics?¹

According to Simon (1956, 1978), an adequate theory of bounded rationality should satisfy two desiderata. First, it should describe the real processes that individuals and institutions use to make actual decisions. Revealing these processes would move economics and psychology beyond “as-if” theories of maximizing expected utility, which in Simon’s view do not even remotely describe the actual processes; see Simon (1978) riposte to Friedman (1953) polemic against realism in economic theory. Second, an empirically grounded theory of bounded rationality should be applicable to situations in which the “conditions for rationality postulated by the model of neoclassical economics are not met” (Simon 1989, p. 377). In particular, it should extend to situations where an agent cannot choose an optimal action, but instead must “satisfice”—that is, choose an option that meets some predetermined aspiration level.

Simon suggested a direction, but did not advance a theory of bounded rationality. Before his death he wrote, “I did not want to give the impression that I thought I had ‘solved’ the problem of creating an empirically grounded theory of economic phenomena. What I was trying to do was to call attention to the need for such a theory” (personal communication; in Gigerenzer 2004, p. 406). Is the heuristics-and-biases program—a project dedicated to mapping out systematic biases and explaining them in terms of either fallible heuristics or minimal modifications of as-if theories—really *not* the sort of theory Simon envisioned?

On the one hand, the empirical findings of the heuristics-and-biases program supplied Simon with ammunition for his foundational attacks on neoclassical theory: “Some of the most dramatic and convincing empirical refutations of the theory [of subjective expected utility] ... reported by D. Kahneman and A. Tversky” (Simon 1979, p. 506) ostensibly suggest that expected utility maximization does not “provide a good prediction—not even a good approximation—of actual behavior” (p. 506). Consequently, some authors have argued that Simon’s work was the intellectual forerunner of the heuristics-and-biases program. The behavioral economist Thaler (1991), for instance, explained that Kahneman and Tversky have shown that “mental illusions should be considered the rule rather than the exception. Systematic, predictable differences between normative models of behavior and actual behavior occur because of what Simson [sic] (1957, p. 198) called ‘bounded rationality’” (p. 4).

¹ There is a third interpretation of bounded rationality, particularly prevalent among neoclassical economists. According to this view, bounded rationality is nothing new. Ultimately, models of bounded rationality are fully optimal procedures that take the costs in terms of time, computation, money, or any other resources being spent into account (e.g., Sargent 1993; Arrow 2004). Simon unapologetically rejected this approach, which he viewed as reductionist: “bounded rationality is not the study of optimization in relation to task environments” (Simon 1991, p. 35).

On the other hand, Simon did not grant that his conception of bounded rationality depicts human decision making to be systematically flawed. For him, “bounded rationality is not irrationality” (Simon 1985, p. 297) and, unlike Kahneman (2003), who postulated an all but insurmountable gap between the canons of rationality and human performance, Simon (1979) proclaimed that “if human decision makers are as rational as their limited computational capabilities and their incomplete information permit them to be, then there will be a close relation between normative and descriptive decision theory” (p. 499).

What, then, is bounded rationality in Simon’s view? Perhaps the most important theoretical analogy he offered is this: “Human rational behavior (and the rational behavior of all physical symbol systems) is shaped by a scissors whose two blades are the structure of the task environments and the computational capabilities of the actor” (Simon 1990, p. 7). In other words, human rationality cannot be understood merely by modeling the mental mechanisms underlying human behavior; it is also necessary to elucidate the relationship between these mechanisms and the environments in which they work. Since the mid-1990s, this *ecological* perspective on bounded rationality has guided an important line of research on judgment and decision making that is closely allied with Simon’s conception of bounded rationality—one that is at odds with the view of bounded rationality as a map of systematic biases.

There are different ways to spell out the ecological perspective on human cognition and decision making. One is in terms of Roger Shepard’s *mirror* conception (1994), according to which key aspects of the environment are internalized in the brain’s neuronal machinery “by natural selection specifically to provide a veridical representation of significant objects and events in the external world” (p. 4; for alternative approaches, see Brunswik 1952 or Dhimi et al. 2004). Simon’s scissors analogy appears to suggest that the mind—which Simon (1969) referred to as an adaptive (“artificial”) system—closely *fits* the environment rather than mirroring it. This fit is achieved in the same way as in any adaptive system; the mind has “responded to the shaping forces of an environment to which it must adapt in order to survive” (Simon 1990, p. 2). Yet this adaption need not come about exclusively, if at all, through the process of natural selection (as in Shepard’s conception). Interpreted in a more inclusive sense, adaptation “may contain large components of conscious intention, as in much human learning and problem solving” (p. 2). In order to highlight Simon’s emphasis on adaptive fit to the environment as a constitutive component of bounded rationality, we included the word “adaptive” in the title of this special issue, and various contributors have made use of the notion of adaptive rationality in their articles.

Simon’s scissors analogy implies that two components are needed to explain the behavior of an adaptive system: (1) models of the relevant environmental (ecological) structures and (2) models of the cognitive processes adapted to these environmental structures. The latter models, however, need to acknowledge a basic truth about the human mind (and, indeed, any artificial system):

Because of the limits on their computing speeds and power, intelligent systems must use approximate methods to handle most tasks. Their rationality is bounded. (Simon 1990, p. 6; his emphasis)

Since the mid-1990s, the research program on simple heuristics (also called *fast and frugal heuristics*) has made progress toward a systematic theory of *ecological rationality*. Specifically, it has identified an ensemble of methods that the human mind may use—its “adaptive toolbox”—to make inferences (e.g., the take-the-best heuristic, the recognition heuristic, the fluency heuristic), choices (e.g., the priority heuristic), and allocations (e.g., in games against nature and games against others; see Gigerenzer et al. 1999; Todd et al. 2012; Hertwig and Herzog 2009; Hertwig et al. 2013). Apart from proposing precise descriptive models of various methods, this program has made two major empirical contributions to the psychology of human judgment and decision making.

First, it has found that methods which respect limited cognitive resources—for instance, by limiting information search, using stopping rules, or applying aspiration levels—can result in more accurate inferences or predictions than can optimizing algorithms (e.g., Gigerenzer and Brighton 2009). This finding challenges what is widely held to be one of the general laws of the human mind: According to the *accuracy–effort tradeoff*, the less information, computation, or time a person uses, the less accurate the person’s judgment will be (e.g., Payne et al. 1993). Second, the simple-heuristics program has shown how environmental structures and circumstances (e.g., high level of uncertainty, limited knowledge of the environment) determine when a heuristic performs better than, say, a Bayesian network or logistic regression (for a review of findings, see **Todd and Brighton** in this special issue).²

The Special Issue

These results and the ensuing debates about problems in the study of bounded rationality were the starting point of the workshop *Finding Foundations for Bounded and Adaptive Rationality* in Berlin and this eponymous special issue.

An important problem addressed in the special issue concerns how to select heuristics from the adaptive toolbox when the environmental structures with which they fit vary over time and space. One aspect of this thorny selection problem is that adaptive heuristics tailored to specific environmental structures may, paradoxically, be reliant on the use of optimal selection strategies to be employed in the “right” environments. **Schurz and Thorn** (in this issue) investigate this question in the context of nonstationary environments by analyzing heuristic strategies, applied on a meta-level, for the selection of strategies. The issue of strategy selection is also the focus of the contribution by **Dana and Davis-Stober** (in this issue). Employing recent advances from research on optimal selection of improper linear models, these authors suggest a prescriptive, computational approach to analyze optimal pairing between environments and heuristics, while also admitting “approximate methods” for strategy selection.

Wellen and Danks (in this issue) consider how the fit between models of heuristics and environments can be achieved through learning (e.g., the learning of cues, cue values, cue validities). They argue that the process of learning an

² Gigerenzer and Sturm (2012) argue that such empirical findings serve to “naturalize” rationality.

environmental representation that is available to a subject at the time of judgment or choice should be taken into account in a theory of bounded and adaptive rationality. They offer one possible formal framework of such a learning process and identify predictions based on it. Learning, in terms of search, is also the focus of **Fu**'s contribution (in this issue). Representing an important tradition of investigating the performance of artificial systems in games like chess—solving a game of chess is computationally unfeasible (a perfect game of chess “calls for the examination of more chess position than there are molecules in the universe”; Simon 1990, p. 6)—Fu demonstrates that the level of intelligence of the system depends on the efficiency of the heuristic search process. Specifically, a boundedly rational system of the level of intelligence hinges on how much search the system can forsake and still reach a given level of performance.

For Simon (1991), one important aspect of any adaptive system is that the system's behavior is adapted to the requirements of specific tasks (chosen from a whole population of possible tasks). For empirical scientists, this raises the question of how to analyze the requirements of a given task, including the specification of rational norms for task performance. **Neth, Sims, and Gray** (in this issue) offer a methodological framework, called *rational task analysis*, to study and set benchmarks for bounded rationality. They expound and discuss rational task analysis by comparing it with related approaches and presenting three informative case studies, each offering insights into studying and understanding human rationality (and the often claimed lack thereof).

As highlighted earlier, the psychological investigation of human decision making has given rise to two paradigmatic research programs over the past five decades, the heuristics-and-biases program and the simple-heuristics program. Both programs understand their theoretical postulates and major results as contributing to a theory of bounded rationality. Moreover, both conceptualizations of bounded rationality have been drawn on to develop policy programs designed to enable people to make better decisions. The more established and widely discussed policy program is the *Nudge* program (with its underlying political philosophy of *libertarian paternalism*; Thaler and Sunstein 2008), which draws on the heuristics-and-biases program. **Grüne-Yanoff and Hertwig** (in this issue) review the *Nudge* approach and a contrasting approach to behavior change, which they call *Boost*, that draws on the simple-heuristics program. Their methodological approach consists in identifying the necessary assumptions underlying each policy approach and studying the extent to which the associated research program is consistent with these assumptions. In other words, they analyze policy–theory coherence, thus identifying strengths and handicaps of both policy programs.

Finally, let us emphasize that it is a great honor for us that the late **Patrick Suppes**, to whom this special issue is dedicated, contributed a paper on the topic of bounded rationality. Pat expressed enthusiasm about contributing to this special issue, undertaking to write about the concept of bounded uncertainty, a subject he believed has been neglected in discussions of bounded rationality. His paper underwent the usual peer-review process, and he made a serious effort to revise it in light of the reviewers' extensive comments. As a testament to his lifelong dedication to his scholarly work, it was a mere 10 days before his death that Pat sent us a

revised manuscript, asking one of us (Pedersen) to make any additional revisions to the manuscript on his behalf.

In his paper, Pat argues that uncertainty, as measured by entropy, can be treated as a fundamental concept akin to qualitative comparative probability and that it is, in a very important sense, bounded in the context of real-world inquiry. He presents a new quantitative representation theorem for uncertainty and focuses on the significance of bounded uncertainty in the design of experiments and analysis of data. In his view, the concept of bounded uncertainty should play a significant role in any general account of bounded rationality.

This special issue gathers together first-rate research on the foundations of bounded and adaptive rationality. Yet it cannot lay claim to being anything more than a status report. It reports on some of the interesting challenges, findings, and advances that have resulted from bringing decision scientists and philosophers together to think through difficult questions raised by the new science of heuristics. We thank all contributors, discussants, and reviewers for their participation in the ongoing endeavor to understand bounded and adaptive rationality.

Acknowledgments We thank the German Research Foundation for grants to Ralph Hertwig as part of the Priority Program *New Frameworks for Rationality* (SPP 1516, HE 2768/7-1 and HE 2768/7-2); we also thank the Priority Program for financially supporting the workshop “Finding Foundations for Bounded and Adaptive Rationality.” We are grateful to the members of the Center for Adaptive Rationality and, especially, to Mauren Antkowsky, Yvonne Bennett, Manuela Meermann, and Renata Suter for helping to organize this workshop. In addition, we thank Susannah Goss and Katja Münz for their editorial assistance with this introduction. Finally, we extend our gratitude to Gregory Wheeler, Editor-in-Chief of *Minds and Machines*, for his extraordinary support at all stages of the process of preparing this special issue.

References

- Arrow, K. J. (2004). Is bounded rationality unboundedly rational? Some ruminations. In M. Augier & J. G. March (Eds.), *Models of a man: Essays in memory of Herbert A. Simon* (pp. 47–55). Cambridge, MA: MIT Press.
- Brunswik, E. (1952). The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1, No. 10, pp. 656–760). Chicago, IL: University of Chicago Press.
- Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature*, 34, 669–700.
- Dhmi, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, 130, 959–988.
- Friedman, M. (1953). *Essays in positive economics*. Chicago, IL: University of Chicago Press.
- Gigerenzer, G. (2004). Striking a blow for sanity in theories of rationality. In M. Augier & J. G. March (Eds.), *Models of a man: Essays in memory of Herbert A. Simon* (pp. 389–409). Cambridge, MA: MIT Press.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1, 107–143.
- Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). *Heuristics: The foundations of adaptive behavior*. New York, NY: Oxford University Press.
- Gigerenzer, G., & Sturm, T. (2012). How (far) can rationality be naturalized? *Synthese*, 187, 243–268.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
- Hertwig, R., & Herzog, S. M. (2009). Fast and frugal heuristics: Tools of social rationality. *Social Cognition*, 27, 661–698.
- Hertwig, R., Hoffrage, U., & the ABC Research Group. (2013). *Simple heuristics in a social world*. New York, NY: Oxford University Press.

- Kahneman, D. (2000). Preface. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames* (pp. ix–xvii). Cambridge, UK: Cambridge University Press.
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, *93*, 1449–1475.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Strauss and Giroux.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.
- Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, *27*, 313–327.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge University Press.
- Sargent, T. J. (1993). *Bounded rationality in macroeconomics*. Oxford, UK: Oxford University Press.
- Shepard, R. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review*, *1*, 2–28.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.
- Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Simon, H. A. (1978). Rationality as process and as product of thought. *American Economic Review*, *68*, 1–16.
- Simon, H. A. (1979). Rational decision making in business organizations. *The American Economic Review*, *69*, 493–513.
- Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.
- Simon, H. A. (1985). Human nature in politics: The dialogue of psychology with political science. *American Political Science Review*, *79*, 293–304.
- Simon, H. A. (1989). The scientist as problem solver. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 375–398). Hillsdale, NJ: Erlbaum.
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1–19.
- Simon, H. A. (1991). Cognitive architectures and rational analysis. Comment. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 25–39). Hillsdale, NJ: Erlbaum.
- Thaler, R. H. (1991). *Quasi rational economics*. New York, NY: Russell Sage Foundation.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
- Todd, P. M., Gigerenzer, G., & the ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.