8.5 Bounded Rationality: A Vision of Rational Choice in the Real World

Ralph Hertwig and Anastasia Kozyreva

Summary

This chapter focuses on the conception of bounded rationality introduced by Herbert Simon as an alternative to the perfect rationality of the omniscient homo economicus, then further developed in psychology and economics. Bounded rationality is a principle stating that real-world rational cognition is limited by bounds on time, computational power, foresight, and knowledge; it is also a theory of what it means to be rational given limited human cognition and an uncertain and complex world. We outline the foundations of bounded rationality in Simon's work and explore interpretations of bounded rationality in three research programs: in economics, the optimization-under-constraints program; in psychology, the heuristics-and-biases program and the program on ecological rationality.

> It is time to take account . . . of the empirical limits on human rationality, of its finiteness in comparison with the complexities of the world with which it must cope. (Simon, 1957b, p. 198)

> Broadly stated, the task is to replace the global rationality of economic man with the kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms . . . in the kinds of environments in which such organisms exist. (Simon, 1955, p. 99)

1. Foundations of Bounded Rationality

1.1 From Metaphors to Models of Choice

If there was any period in human history during which the Enlightenment ideals of rationality and humanity were at risk of being shattered completely, it was the 20th century. Paradoxically, this century, abounding with the horrors of human irrationality and unprecedented self-destruction, was also the most fruitful period for advancing both formal normative theories of rational choice and psychologically realistic models of reasonable behavior. Herbert Simon's work on bounded rationality was one such 20th-century advancement. It established a new requirement for theorizing about human choice: the need to understand how real people, not agents with supreme foresight and unlimited computational skills, make decisions when faced with limited time and knowledge. The time was ripe to comprehend how people achieve their goals in the face of a staggering environmental complexity that surpasses the limits of the computational resources of any human being or any information-processing system.

Simon introduced the concept of bounded rationality in the mid-1950s to describe a new theoretical and empirical approach to studies of choice and rational behavior. First, he defined bounded rationality as a principle that emphasizes the limits on computational capacities of agents in the real world who have only limited time, knowledge, foresight, and cognitive resources:

> The principle of bounded rationality: The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world—or even for a reasonable approximation to such objective rationality. (Simon, 1957b, p. 198)

He thereby positioned bounded rationality as an alternative to the concepts of perfect or substantive rationality rooted in classical rational choice theory that were prominent in the neoclassical economics of his day (more on this juxtaposition in section 1.2). The idea of cognitive limitations and impractical normative standards in the real world was already present in his early research on decision making within organizations (Simon, 1947/1957a). Although Simon initially embraced the normative standards of rational choice theory, showing that real people fall short of these standards,[1] in later works, he progressed toward a positive theory of bounded rationality (most notably beginning with Simon, 1955, 1956). This extension from principle to positive theory implies that bounded rationality is not merely a simplification of perfect or substantive rationality but a descriptive and prescriptive model of actual choice behavior in the real world. In fact, Simon's ultimate endeavor was to reconstruct and reinvent the "theory of the rational" (Simon, 1957b, p. 200). On this view, bounded rationality represents a new vision of what it means to be rational given both the inescapable uncertainty and complexity of the world and the limitations of the human cognitive system. Both the principle of bounded rationality and the vision of bounded rationality as a new descriptive and prescriptive theory of rational choice imply that understanding (and eventually predicting) the choice behavior of real humans is only possible when psychological processes are empirically analyzed. This study of what Simon called "procedural rationality" ultimately amounts to a theory of the efficient computational procedures by which agents (individual human beings, organizations, or computers) arrive at good solutions to the problems they face (Simon, 1976).

Simon went on to develop his approach to bounded rationality across a diverse set of scientific disciplines such as economics, cognitive psychology, and artificial intelligence, adding new levels to his concept over time. One of the most important developments concerned his inquiry into an adaptive, or ecological, level of rationality. Specifically, he stressed that the essence of rational behavior consists in how an organism can adapt in order to achieve its goals under the constraints of its environment and its own cognitive limitations. These two dimensions of rational behavior—cognitive and environmental—gave rise to the scissors metaphor that encapsulates bounded rationality's theoretical core: "Human rational behavior (and the rational behavior of all physical symbol systems) is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1990, p. 7).

The scissors metaphor was not the only one Simon used to portray human choice and choice environments. As he recalled in his autobiography, he saw the labyrinth as another powerful symbol for decision making. Simon (1991) envisioned decision making "in terms of successive choices along a branching path" (p. 86) rather than as an optimizing process. The first step in his search to turn his metaphors of bounded rationality into realistic models of rational behavior was to criticize the narrow and unrealistic conception of rationality found in neoclassical economics.

1.2 The Limits of the Olympian Model of Rationality

Simon (1983) opposed what he called the "Olympian model" of rationality (p. 19), which encompasses concepts such as perfect or substantive rationality, (subjective) expected utility theory, and the unfailingly rational homo economicus.[2] In order to satisfy its normative demands, this model requires a decision maker with unlimited cognitive resources who can form mental models representing, in a collectively exhaustive and mutually exclusive way, all future relevant states of the world. Simon argued that while rational choice theory was worthy of a home in Platonic heaven, it was simply out of place in the real world (Simon, 1983, p. 13). In "A Behavioral Model of Rational Choice" (Simon, 1955), he spelled out the unrealistic assumptions of perfectness and logical omniscience that the Olympian approach to rationality made about the agent:

> If we examine closely the "classical" concepts of rationality outlined above, we see immediately what severe demands they make upon the choosing organism. The organism must be able to attach definite pay-offs (or at least a definite range of pay-offs) to each possible outcome. This, of course, involves also the ability to specify the exact nature of outcomes—there is no room in the scheme for "unanticipated consequences." The pay-offs must be completely ordered—it must always be possible to specify, in a consistent way, that one outcome is better than, as good as, or worse than any other. And, if the certainty or probabilistic rules are employed, either the outcomes of particular alternatives must be known with certainty, or at least it must be possible to attach definite probabilities to outcomes. (pp. 103–104)

In actual human choice, such demanding requirements can rarely be met. Outside what Savage (1954) called "small worlds"—highly simplified environments such as monetary gambles—real people cannot live up to the ideal of making decisions by specifying all possible outcomes, assigning probabilities and values to each, and then maximizing the expected payoffs. It would be a misrepresentation, however, to assume that people cannot live up to this decision-making ideal due to cognitive destitution or a mere lack of skill; Simon (1972) credited the *irreducible uncertainty* inherent in any human reckoning about the future as the reason people did not act according to the Olympian model.

Consider firefighters. Like emergency room doctors, police officers, and soldiers, firefighters face enormous time pressure, high stakes, and inescapable uncertainty. They do not have the luxury of mentally generating all possible courses of actions, specifying their respective outcomes, and then evaluating them. Not only do firefighters face epistemic uncertainty, but they also face the aleatory uncertainty[3] inherent in their environment. The Olympian model thus seems to be an obscenely lofty conception of decision making. Professional experts do not make decisions by comparing all possible options but rather by generating a single good course of action from the start (Klein, 1998, p. 17). By drawing on a repertoire of past situations compiled from real and virtual experience, firefighters and their ilk quickly identify a plausible course of action and play it out mentally to test whether it could work in the current situation. If success seems plausible, it will be implemented; otherwise, it will be modified or replaced by the next-best option until a "good enough" alternative is found. This is an example of a fast and intuitive problem-solving performance—one for which the expert cannot necessarily describe in detail the underlying reasoning processes. Moreover, the key to experts' success lies in the way they approach and analyze their environments. Returning to Simon's scissors, one blade—the environment—provides the cues that are in turn matched by the second blade—the mind—which draws on its resources (e.g., memory) to find a suitable decision strategy. On this view, "intuition is nothing more and nothing less than recognition" (Simon, 1992, p. 155).

Next to irreducible uncertainty and time pressure, there is another key reason that the Olympian model prospers in Platonic heaven but withers in earthly reality: computational intractability. Consider chess, a game Simon studied extensively (Chase & Simon, 1973). Chess offers a choice set of about 30 legal moves (a number that stays more or less constant until the end of the game) and a time horizon of about 40 moves per player until one party concedes; 30^{80} possible sequences (about 10^{118}) follow from the original position (see also Shannon, 1950). No human or computer can generate and evaluate all these consequences. Even Deep Blue, the IBM computer that beat chess world champion Garry Kasparov in 1997 and that could examine some 300 million possible moves per second, had to evaluate positions through "a high-performance alpha-beta search that expands a vast search tree by using a large number of clever heuristics and domain-specific adaptations" (Silver et al., 2018, p. 1). Relative to many real-world decisions, however, chess is a piece of cake; indeed, "the problem space associated with the game of chess is very much smaller than the space associated with the game of life" (Simon, 1976, p. 72). In many social interactions, the rules are not as well defined as they are in chess, the set of possible actions is much vaster, there are more than two parties involved, and goals are unclear or, even worse, in conflict, thus compounding the intractability problem with additional complexity. If chess is computationally intractable and thus beyond the reach of Olympian optimization, so are many of the decisions people face in the real world.

In other words, even though the classical understanding of rationality as set out in Olympian models is theoretically coherent and even appealing, it profoundly misrepresents the reality of decision making: the set of decisions where Platonic heaven and earthly reality intersect is very narrow.[4] For Simon (1989), the fundamental question for the study of bounded rationality was, therefore, "How do human beings reason when the conditions for rationality postulated by the model of neoclassical economics are not met?" (p. 377). This question has an important implication. Classical concepts of rationality not only fail to describe what people actually do but also fail to offer procedural advice about how to find solutions to many real-world problems. The study of boundedly rational decision making, on the other hand, is ultimately concerned with both descriptive and prescriptive questions—a point to which we return later.

1.3 Principles of Bounded Rationality

Firefighters face limits on time, information, and certainty—yet they still make decisions, as do emergency room doctors and chess players. How do reasonable people decide when optimization is out of reach, and how can scientists examine this behavior? In order to show how Simon answers this question, we distinguish four principles underlying his investigation of bounded rationality.

The first principle calls for a behaviorally informed theory of choice, which should take into account and specify what conditions govern human behavior in different real-world situations. In Simon's (1992) words, "The study of the behavior of an adaptive system like the human mind is not a logical study of optimization but an empirical study of the side conditions that place limits on the approach to the optimum" (p. 157). Failing to understand not just the "design product (the alternative finally chosen) but the design process as well" (p. 156) compromises behavioral science's ability to explain and predict behavior. This postulate was a bone of contention between Simon and Milton Friedman, who fiercely defended classical economic theory by arguing that "'complete realism' is clearly unattainable, and the question whether a theory is realistic 'enough' can be settled only by seeing whether it yields predictions that are good enough for the purpose at hand" (Friedman, 1953, p. 41). Simon, however, insisted that a realistic account of the cognitive processes was required to successfully predict human behavior (see also Simon, 1976, on procedural rationality).

A second principle is that realistic models will be models of "approximate methods" (Simon, 1990, p. 6). "Approximate" here should not be interpreted to mean that these methods will produce inferior solutions or lack cognitive sophistication. Simple methods, for instance, can exploit evolved cognitive, visual, and motoric capacities that can be complex and demanding but are nevertheless easy for the mind to execute, such as pattern recognition, emotions (see Hanoch, 2002), tracking motion in space, and the ability to adaptively forget (Schooler & Hertwig, 2005). By exploiting the mind's sophisticated capacities, approximate methods can remain computationally slim and operate under constraints on time and information.

A third principle of bounded rationality is that when optimization is out of reach, people *satisfice*—they look for good-enough answers. Satisficing does not mean merely taking a good-enough alternative and giving up on the best possible alternative. Rather, satisficing is "the process of finding alternatives by heuristic search with the use of a stop rule based upon adjustable aspirations" (Simon, 1982, p. 323). In natural situations, alternatives are often encountered sequentially (e.g., potential mating partners; Miller & Todd, 1998), and the number of available alternatives is almost always too large to be exhaustively explored; furthermore, alternatives may vanish if not chosen immediately. Under these circumstances, a limited search is implemented and the point at which a search will be terminated is determined by an aspiration level—"a goal variable that must be reached or surpassed by a satisfactory decision alternative" (Selten, 2001, pp. 13–14). Aspiration levels can be adjusted depending on one's success in finding satisfying options. Note that not every boundedly rational behavior requires satisficing. Rather, a range of heuristic processes can be employed to reach good decisions in both strategic games and games against nature (Hertwig & Herzog, 2009; see also section 2.3 of this chapter).

A final principle of the empirical study of bounded rationality requires taking into account the properties of an organism's environment. In *The Sciences of the Artificial* (1969/1996), Simon argued that an organism's behavior results not merely from its inherent characteristics but also from the structures of its surroundings. Taking the path of an ant on a beach as an example, Simon claimed that an "ant, viewed as a behavior system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment" (p. 52). An ant's environment may be an obstacle course, forcing the ant to repeatedly change direction around pebbles, rocks, and puddles (figure 8.5.1). An observer who saw only the ant's laborious path, without the properties of the environment, may be tempted to attribute substantial complexity to the cognitive mechanisms underlying the ant's behavior. Yet, if the ant's behavior is understood as the result of its interaction with its environment, the explanation may be much more mundane. The mechanisms producing the complex path may be simple rules such as, "If the left vision field sensors for obstacles lights up, turn to the right, and vice versa." Even though actual ants might be guided by a multitude of other factors as well, this example reveals that the study of bounded rationality must include the study of the structural properties of the surroundings in which organisms, whether ants or humans, make decisions. Explaining complex behaviors exclusively in cognitive terms risks misattributing—or worse, profoundly misconstruing—the causes of the complexity.

Having clarified the basic principles behind Simon's view, we now turn to various research programs in psychology and economics that claim to have further developed the theory of bounded rationality. As we will see, these programs are quite diverse and give varying weight to the theoretical and methodological notions and principles reviewed above.

2. Interpretations of Bounded Rationality

There are at least three main research programs aiming to interpret bounded rationality and develop it further (see also Gigerenzer & Selten, 2001): the economic research program of *optimization under constraints*, the approach in psychology known as the *heuristics-and-biases program* that led to the foundation of behavioral economics (and behavioral law), and a further program in psychology known as *ecological rationality* or the study of *fast-and-frugal heuristics*.

2.1 Optimization under Constraints

Simon's candid and fundamental criticism of neoclassical economics and its view of global and omniscient

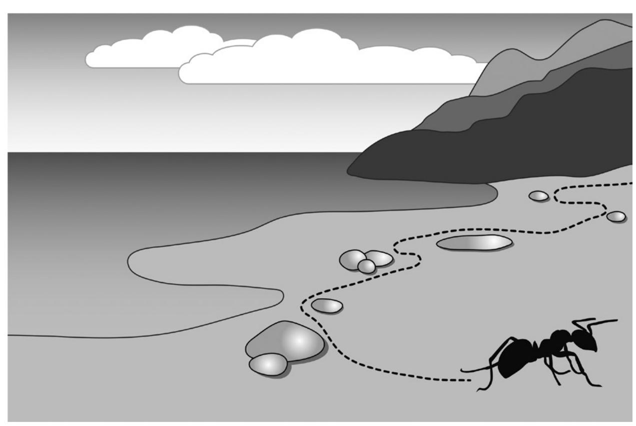

Figure 8.5.1
An ant walking on a beach illustrates that a behavior (here, the path the ant takes) that might appear complex to an external observer is actually a function of the interaction between the organism and its environment.

rationality was generally not met with enthusiasm by economists. However, in an attempt to reconcile neoclassical economics with behavioral and cognitive limits, George Stigler, a leading figure of the Chicago school of economics, proposed a notion often described as "optimization under constraints" (Stigler, 1961).[5] As a former student of Frank Knight, Stigler (1961) was mindful of the role of uncertainty in human choice, stating that "our understanding of economic life will be incomplete if we do not systematically take account of the cold winds of ignorance" (p. 224).

In order to incorporate ignorance—or, to use a less pejorative term, incomplete knowledge—in the process of choice, Stigler (1961) focused on one key idea of bounded rationality: the concept of limited search (and, by extension, the respect for real decision makers' finite resources). The method for determining when to stop searching is interpreted in terms of a cost–benefit analysis. On this view, conditions of limited search and less-than-perfect information appear to easily square with models of optimization. Specifically, the models in this class assume that a stopping rule optimizes with respect to the relevant currency (e.g., time, computation, money). Thus, when the cost of searching for new alternatives or other pieces of information exceeds the benefits of further exploration, the search will be terminated. For example, when looking for a new car, a buyer explores the market until the expected marginal costs (i.e., time and effort) of looking for new options exceed the expected marginal benefits (i.e., finding a car that more closely meets the buyer's preferences). Although this conception of psychologically realistic decision making apparently reconciles bounded rationality with optimization (see also Sargent, 1993), the cost–benefit analysis imposes the "burden of estimating the expected marginal return of search and the opportunity cost" (Simon, 1982, p. 296). Indeed, optimization under constraints requires the same amount of information and even more knowledge and computation (e.g., determining the costs of continuing search, such as search's opportunity costs; Vriend, 1996) compared to neoclassical optimization models. Moreover, as Simon (1979) noted, it "poured the search theory back into the old bottle of utility maximization" (p. 503) without paying due attention to a key element of the theory of bounded rationality, namely, that choice can be made with incomplete information and without carrying out optimization procedures.

2.2 Heuristics and Biases

The theory of bounded rationality is meant to be a behaviorally and psychologically grounded approach to decision making. One of the research programs that has contributed significantly to developing this approach is the work instigated by Amos Tversky and Daniel Kahneman. Their heuristics-and-biases approach has mapped the impact of cognitive limitations on people's judgments and decisions and documented a large catalog of systematic deviations from norms of rationality, drawn from probability theory, statistics, and axioms of rational choice (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). Kahneman interpreted bounded rationality in terms of behavior that diverges from such norms and, by extension, optimality. In this spirit, he claimed, "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" (Kahneman, 2003, p. 1449).

This statement both highlights the key normative assumption of the heuristics-and-biases program and outlines the contours of its map of bounded rationality. Rational-agent models (Olympian models, according to Simon) as well as their axioms and assumptions (e.g., Bayesian probability updating) are explicitly acknowledged as the benchmarks against which people's judgments should be compared. This means that the decision problems in question are ones in which an optimal or normative solution exists. By extension, bounded rationality is not so much a map of human choice in the world of the unknown and the uncertain as it is an empirical collection of decisional deficiencies in problems with assumed normative solutions. This is also evident in prospect theory, the most influential descriptive alternative choice theory to expected utility theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Indeed, as Kahneman emphasized, their aim was to "assemble the minimal set of modifications of expected utility theory that would provide a descriptive account of . . . choices between simple monetary gambles" (Kahneman, 2000, p. x). The theory's goal was to explain a wide range of violations of axioms and predictions of expected utility theory, thus accepting the normative force of the theory. Prospect theory enriches the expected utility framework by introducing a few psychological concepts (e.g., nonlinear probability weighting, reference points, loss aversion; for a detailed account of prospect theory, see chapter 8.3 by Glöckner, this handbook). Yet its scope, at least in its original form, is confined to the world of risk (Savage's small worlds) with known possible outcomes and without the possibility of surprises. As Simon repeatedly stressed, these are not the conditions under which individuals and organizations make decisions most of the time.

Outside of the domain of risky choice, the heuristics-and-biases program has invoked a small set of heuristics, or mental shortcuts, to explain the myriad errors—also called "cognitive illusions"—that people commit in their inferences and decisions. Specifically, the program's argument is that "people rely on a limited number of heuristic principles which reduce the complex task of assessing probabilities and predicting values to simpler judgmental operations. In general these heuristics are quite useful, but sometimes they lead to severe and systematic errors" (Tversky & Kahneman, 1974, p. 1124).

Three key heuristics—representativeness, availability, and anchoring and adjustment—have been proposed to account for a wide range of cognitive illusions such as the base-rate fallacy, the overconfidence bias, the conjunction fallacy, the hot-hand fallacy, and the failure to appreciate regression to the mean or the law of large numbers. Two theoretical premises buttress these heuristics. One is that of attribute substitution, such as similarity-for-probability substitution, which is the mechanism behind the representativeness heuristic (Kahneman & Frederick, 2002). It describes the idea that a heuristic evaluates a target attribute of a judgment object (e.g., the likelihood that *X* is a *Y* or a *Z*) by substituting it with a property of the object that comes more readily to mind (e.g., the degree to which *X* resembles a *Y* or a *Z*). This occurs when a person who learns that Linda is an outspoken woman concerned with social justice judges that it is more likely that Linda is a "feminist bank teller" than—as is necessarily at least as likely—a "bank teller," because they believe Linda's characteristics are representative of a feminist (Tversky & Kahneman, 1983). Error-prone attribute substitution occurs when three conditions are satisfied: first, the target attribute (e.g., probability) is cognitively relatively inaccessible; second, the substituting attribute (e.g., similarity) is highly accessible; and, third, the mind's "controller" does not reject the substitution.

This controlling entity relates to the second theoretical premise key to Kahneman's (2011) view of bounded rationality—namely, that two, admittedly fictitious, cognitive systems make up the human mind. System 1 is automatic, effortless, and associative. Reasoning errors are typically the product of this speedy system and its propensity to propose the solution that most readily comes to mind. The awareness that an error may have occurred requires the more reflective and effortful System 2. This system acts like a teacher identifying a student's mistake and rectifying the process that produced it. But System 2 also tires easily. Too often, instead of analyzing the proposals of System 1, System 2 is content with the solutions offered, thus leaving decision makers unaware that they are about to commit an error.

The heuristics-and-biases interpretation of bounded rationality in terms of a map of the systematic deviations between people's behavior and norms of rationality has been highly influential within psychology and its neighboring fields and has even prompted new fields of research such as behavioral economics (e.g., Thaler, 2016) and behavioral law (e.g., Sunstein, 2000). However, it has also provoked a number of critical objections. One is that the standards against which human reasoning is compared in this program are narrow and predominantly content-free coherence norms (Arkes, Gigerenzer, & Hertwig, 2016). Another is that the identified cognitive illusions are by no means as impervious to change (e.g., via different formats for representing probabilistic information; see, e.g., Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000) as has often been claimed. Finally, it has been demonstrated that at least some behaviors seemingly based on cognitive illusions are not best understood as manifestations of human irrationality; instead, they may be the product of evolved and adaptive decision rules. These rules, however, are not adapted to experimentalists' simplified environments and so are wrongly interpreted as irrational behavior (Fawcett et al., 2014; see also chapter 10.6 by Cosmides & Tooby, this handbook, for an evolutionary perspective on human reasoning).

Interestingly, this is a development already predicted by Simon (1957b), who wrote, "I believe that the return swing of the pendulum will begin, that we will begin to interpret as rational and reasonable many facets of human behavior that we now explain in terms of affect" (p. 200).

2.3 Ecological Rationality

The third influential interpretation of bounded rationality is ecological rationality and the study of fast-and-frugal heuristics.[6] This approach aims to understand boundedly rational decision mechanisms based on how they match the statistical structures of choice environments. Following Simon's emphasis on understanding and modeling rational behavior as shaped by both cognitive and environmental structures, research on ecological rationality explores the adaptive toolbox of simple heuristics that human minds have developed through individual, cultural, or evolutionary learning (see Gigerenzer, Todd, & ABC Research Group, 1999; Hertwig, Hoffrage, & ABC Research Group, 2013; Hertwig, Pleskac, Pachur, & the Center for Adaptive Rationality, 2019; Todd, Gigerenzer, & ABC Research Group, 2012). Simple heuristics, understood as the boundedly rational mind's main tools in games against nature and social games (Hertwig & Herzog, 2009), are defined as follows: "A heuristic is a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods" (Gigerenzer & Gaissmaier, 2011, p. 454).

Proposing a new "theory of the rational" (Simon, 1957b, p. 200), ecological rationality strives to model rational behavior in terms of correspondence norms (Hammond, 2000). Correspondence norms represent a measure of cognitive success in achieving one's goals in the world—a consequentialist interpretation of rationality in cognition (Kozyreva & Hertwig, 2021; Schurz & Hertwig, 2019; see also chapter 1.3 by Schurz, this handbook). Depending on the environment, cognitive success could be measured in terms of criteria such as accuracy, speed, frugality, robustness, or accountability. An important methodological principle in research on ecological rationality is comparatively testing heuristic strategies against informationally and computationally complex strategies. Ideally, the success of each strategy should be measured against that of others and across a reference class of environments (for examples of such comparative tests, see Gigerenzer & Brighton, 2009; Spiliopoulos & Hertwig, 2020). An important insight stemming from tests involving a set of strategies and environments is that heuristics do not result in good or bad performance in and of themselves; a heuristic's performance is relative to the structure of the environment in which it is employed. If a heuristic is able to exploit the structure at hand, the heuristic can be surprisingly accurate. Similarly, it can yield dismal performance when facing environment structures that do not match its architecture or its assumptions about the environment. Just as the ant's complex path cannot be understood apart from the environment in which it unfolded, cognitive tools and their ecological rationality cannot be divorced from the environmental structures they may exploit.

Research on ecological rationality has produced three key discoveries pertaining to the understanding of bounded rationality. First, bounded rationality does not mean that simplicity in search, information integration, and decision rules will inevitably result in second-rate performance. In fact, less is often more. Second, bounded rationality requires an ongoing examination of the environmental structures that support heuristics' performance. To "describe, predict and explain the behavior of a system of bounded rationality, we must both construct a theory of the system's processes and describe the environments to which it is adapting" (Simon, 1990, p. 7). One could go even further. Rather than cataloging piecemeal descriptions of environments, one could harness existing theories of environments that shed light on the factors and dynamics that amplify, attenuate, or even wipe out the environmental structure in question. Based on such a theory, one would then examine how the environmental processes interact with theories of the system's process—for an illustration of this approach, see an analysis of the risk–reward structure using the foraging theory of ideal free distribution in Pleskac, Hertwig, Leuker, and Conradt (2019). Third, heuristics can enable good performance in a wide range of domains (e.g., judgment, choice, inference, classification) in which people face uncertainty, lack of information, and time pressure.

In briefly reviewing the last discovery relating to the often surprising success of heuristics, we start with the question of why people recruit boundedly rational heuristics to begin with. The classical explanation is that people resort to heuristics in a kind of compromise, saving cognitive effort at the cost of accuracy (Shah & Oppenheimer, 2008). Humans and other animals rely on heuristics because searching for and processing information can be taxing, and heuristics offer relief by trading reduced accuracy for faster, more frugal cognition. This accuracy–effort trade-off—variously conceptualized as a rational and beneficial trade-off (Payne, Bettman, & Johnson, 1993) or as one that is constrained by capacity limitations and paid for in terms of reasoning mistakes (the heuristics-and-biases view)—has been touted as a potentially universal law of cognition. But there is an alternative explanation for why people employ heuristics: less computation and less information can result in performance that is as good as, or even better than, that of computationally and informationally complex strategies. In a wide range of studies, often involving computer simulations and analytical work, the ecological rationality program has demonstrated numerous instances of such less-is-more effects (Gigerenzer, Hertwig, & Pachur, 2011; Spiliopoulos & Hertwig, 2020).

One way to understand these effects is through the distinction between prediction and fitting. *Prediction* occurs when the data have not yet been observed and a model, whether heuristic or complex in nature, with fixed parameter values is employed to predict future events. *Fitting*, in contrast, occurs when the data have already been observed and the parameters of a model are estimated in order to maximize the fit between the data and the model's behavior. Generally, the more free parameters a model has, the better its fit. This rule, however, does not hold for predictions. When parameters need to be estimated from small or unreliable samples, the function between predictive accuracy and model flexibility (e.g., the number of free parameters) is typically inversely U-shaped. This means that models that are too simple (e.g., have too few parameters) or too complex can fail in prediction (Pitt, Myung, & Zhang, 2002). Heuristics—at least some—can occupy the sweet spot between too little and too much complexity.

The distinction between prediction and fitting is related to another theoretical construct relevant to understanding why heuristics can perform so well: the bias–variance framework (exploited in machine learning; Geman, Bienenstock, & Doursat, 1992). Heuristics can succeed because they smartly trade off two components of prediction error: *bias* (how well, on average, the model can agree with the ground truth) and *variance* (the variation around this average). Heuristics tend to have higher bias but lower variance than more complex models (with more adjustable parameters), which explains why heuristics perform relatively better when making predictions from small environmental samples (Gigerenzer & Brighton, 2009; Katsikopoulos, Schooler, & Hertwig, 2010). Variance is a substantial source of error when information about the environment is sparse, thus exposing more complex models to the risk of overfitting by virtue of the flexibility granted by their parameters. Heuristics, however, are less flexible and thus less likely to overfit, which gives them the chance to outperform more complex models when knowledge about the environment is incomplete and uncertainty is high.

Still another approach to understanding heuristics' ecological rationality is in terms of the match between environmental structures (i.e., statistical properties that reflect patterns of information distribution in the given ecology) and cognitive strategies (Hogarth & Karelaia, 2006, 2007). For instance, one notable property of an environment is the presence of a *noncompensatory* cue. This is an attribute that has a much higher correlation with the ground truth than all other attributes combined, which means it cannot be outweighed (compensated) by any combination of less valid cues or attributes (for details and other environmental structures, see Şimşek, 2013). Noncompensatory cues can be exploited by lexicographic heuristics such as Take-the-Best (Gigerenzer & Goldstein, 1996) and LEX (Payne et al., 1993) that order attributes according to importance and process them sequentially.

Finally, another significant contribution of ecological rationality explores the ways in which heuristics can be engineered to foster better decisions in uncertain circumstances. As we have seen, the heuristics-and-biases program has focused on a small set of heuristics and refrained from proposing heuristics to help people make better decisions. This is consistent with the view that heuristics cannot compete with, let alone outperform, more complex models. In contrast, the ecological rationality program has suggested and inspired new heuristics (often in terms of fast-and-frugal trees; Martignon, Katsikopoulos, & Woike, 2008) for a wide range of professional uses, including geographic profiling, prescribing antibiotics, and allocating marketing resources (see Gigerenzer et al., 2011). This ongoing work suggests that the study of boundedly rational heuristics not only examines which heuristics people use but also addresses profound prescriptive concerns, thereby providing the basis for boosting people's competences by giving them smart and easy-to-use tools and decision aids (Hertwig & Grüne-Yanoff, 2017; the boosting approach to public policy making is an alternative to the nudging approach that is rooted in the heuristics-and-biases program; Thaler & Sunstein, 2008).

3. Conclusion

Having already inspired groundbreaking investigations into a map of cognitive illusions and an adaptive toolbox of ecologically rational heuristics, the exploration of bounded rationality has nevertheless in many regards only just begun. Great challenges are ahead, including constructing an encompassing theory-based taxonomy of environmental structures that decision makers perceive and the heuristics that exploit them. Another largely neglected research topic is how simple heuristics, by interacting with environment structures as well as with the heuristics of other decision makers, give rise to complexity in the environment (see, e.g., Hertwig, Davis, & Sulloway, 2002; for a related concern, see Schelling, 2006). In the years to come, the study of bounded rationality should further explore and detail the content of the adaptive toolbox beyond simple heuristics, investigating other cognitive tools that people—both as individuals and in groups—use to handle uncertainty (see Hertwig et al., 2019). Perhaps the most ambitious challenge of all is to uncover the extent to which integrating the sciences and methodologies of heuristics, search, and learning (Hertwig, 2015; Lejarraga & Hertwig, in press) can offer a unified framework for the study of the boundedly rational mind. In a future of vast complexity, informational affluence, and potential for surprise, the need for a realistic vision of rational choice will only grow—and with it the scientific exploration of bounded rationality.

Acknowledgments

We are grateful to Max Albert, Hartmut Kliemt, and Gregory Wheeler for their comments and suggestions. We thank Deb Ain for editing the manuscript.

Notes

1. In the second edition of *Administrative Behavior*, Simon (1947/1957) wrote, "Administrative theory is peculiarly the theory of intended and bounded rationality—of the behavior of human beings who *satisfice* because they have not the wits to *maximize*" (p. xxiv). He later asserted, "In *Administrative Behavior*, bounded rationality is largely characterized as a residual category—rationality is bounded when it falls short of omniscience" (Simon, 1979, p. 502).

2. Simon employed several terms throughout his writings to define this theory, including "global rationality," "perfect rationality," "objective rationality," "substantive rationality," "full rationality," "homo economicus," and "Olympian model of rationality." The target of these descriptions was the dominant classical rational choice theory, including, most notably, expected utility and subjective expected utility theories. In rational choice theory, norms of coherence and axioms of expected utility theory (e.g., transitivity, completeness) are taken as benchmarks of rational decision making. This approach, sometimes described as the standard picture of rationality (Stein, 1996), remains the most commonly used normative approach to rationality in decision sciences.

3. The distinction between aleatory and epistemic uncertainty (or objective and subjective uncertainty) stems from the duality inherent in the modern concept of probability, which encompasses both epistemic probability (subjective degrees of belief) and aleatory probability (stable frequencies displayed by chance devices; Hacking, 1975/2006). In a similar vein, epistemic uncertainty refers to incomplete knowledge or information, whereas aleatory uncertainty stems from the statistical properties of the environment, which exist independently of a person's knowledge (Kozyreva & Hertwig, 2019).

4. In his Nobel Prize lecture, Simon (1979) stated, "A strong positive case for replacing the classical theory by a model of bounded rationality begins to emerge when we examine situations involving decision making under uncertainty and imperfect competition. These situations the classical theory was never designed to handle, and has never handled satisfactorily" (p. 497).

5. The idea of optimization under constraints was originally outlined by Good in 1952 when he observed that agents in the real world seek to minimize costs involved in obtaining information and making decisions (in terms of both time and effort; see also Wheeler, 2018/2019).

6. This use of the term "ecological rationality" is not to be confused with the terminologically similar but conceptually different concept developed by Vernon Smith (2003).

References

Arkes, H. R., Gigerenzer, G., & Hertwig, R. (2016). How bad is incoherence? *Decision*, *3*(1), 20–39.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

Fawcett, T. W., Fallenstein, B., Higginson, A. D., Houston, A. I., Mallpress, D. E., Trimmer, P. C., & McNamara, J. M. (2014). The evolution of decision rules in complex environments. *Trends in Cognitive Sciences*, *18*(3), 153–161.

Friedman, M. (1953). *Essays in positive economics*. Chicago, IL: University of Chicago Press.

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

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(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*(4), 650–669.

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. Oxford, England: Oxford University Press.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

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

Good, I. J. (1992). Rational decisions. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in statistics: Foundations and basic theory* (pp. 365–377). New York, NY: Springer. (Original work published 1952)

Hacking, I. (2006). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference* (2nd ed.). Cambridge, England: Cambridge University Press. (Original work published 1975)

Hammond, K. R. (2000). Coherence and correspondence theories in judgment and decision making. In T. Connolly, H. A. Arkes, & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 53–65). Cambridge, England: Cambridge University Press.

Hanoch, Y. (2002). "Neither an angel nor an ant": Emotion as an aid to bounded rationality. *Journal of Economic Psychology*, *23*(1), 1–25.

Hertwig, R. (2015). Decisions from experience. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1, pp. 240–267). Chichester, England: Wiley Blackwell.

Hertwig, R., Davis, J. N., & Sulloway, F. J. (2002). Parental investment: How an equity motive can produce inequality. *Psychological Bulletin*, *128*(5), 728–745.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), 973–986.

Hertwig, R., & Herzog, S. M. (2009). Fast and frugal heuristics: Tools of social rationality. *Social Cognition*, *27*(5), 661–698.

Hertwig, R., Hoffrage, U., & ABC Research Group. (2013). *Simple heuristics in a social world*. Oxford, England: Oxford University Press.

Hertwig, R., Pleskac, T. J., Pachur, T., & Center for Adaptive Rationality. (2019). *Taming uncertainty*. Cambridge, MA: MIT Press.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*(5500), 2261–2262.

Hogarth, R. M., & Karelaia, N. (2006). Regions of rationality: Maps for bounded agents. *Decision Analysis*, *3*(3), 124–144.

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*(3), 733–758.

Kahneman, D. (2000). Preface. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames* (pp. ix–xvii). Cambridge, England: Cambridge University Press.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475.

Kahneman, D. (2011). *Thinking, fast and slow*. London, England: Penguin Books.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge, England: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*(2), 263–291.

Katsikopoulos, K. V., Schooler, L. J., & Hertwig, R. (2010). The robust beauty of ordinary information. *Psychological Review*, *117*(4), 1259–1266.

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Kozyreva, A., & Hertwig, R. (2021). The interpretation of uncertainty in ecological rationality. *Synthese*, *198*, 1517–1547. doi:10.1007/s11229-019-02140-w

Lejarraga, T., & Hertwig, R. (in press). How experimental methods shaped views on human competence and rationality. *Psychological Bulletin*.

Martignon, L., Katsikopoulos, K. V., & Woike, J. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology*, *52*(6), 352–361.

Miller, G. F., & Todd, P. M. (1998). Mate choice turns cognitive. *Trends in Cognitive Sciences*, *2*(5), 190–198.

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

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*(3), 472–491.

Pleskac, T. J., Hertwig, R., Leuker, C., & Conradt, L. (2019). Using risk–reward structures to reckon with uncertainty. In R. Hertwig, T. Pleskac, T. Pachur, & the Center for Adaptive Rationality (Eds.), *Taming uncertainty* (pp. 51–70). Cambridge, MA: MIT Press.

Sargent, T. J. (1993). *Bounded rationality in macroeconomics: The Arne Ryde memorial lectures*. Oxford, England: Clarendon Press.

Savage, L. J. (1954). *The foundations of statistics*. New York, NY: Wiley.

Schelling, T. C. (2006). *Micromotives and macrobehavior*. New York, NY: Norton.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, *112*(3), 610–628.

Schurz, G., & Hertwig, R. (2019). Cognitive success: A consequentialist account of rationality in cognition. *Topics in Cognitive Science*, *11*(1), 7–36.

Selten, R. (2001). What is bounded rationality? In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 13–36). Cambridge, MA: MIT Press.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222.

Shannon, C. E. (1950). Programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *41*(314), 256–275.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., . . . Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140–1144.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.

Simon, H. A. (1957a). *Administrative behavior* (2nd ed.). New York, NY: Macmillan. (Original work published 1947)

Simon, H. A. (1957b). *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting*. New York, NY: Wiley.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, *1*(1), 161–176.

Simon, H. A. (1976). From substantive to procedural rationality. In T. J. Kastelein, S. K. Kuipers, W. A. Nijenhuis, & G. R. Wagenaar (Eds.), *25 years of economic theory: Retrospect and prospect* (pp. 65–86). Boston, MA: Springer.

Simon, H. A. (1979). Rational decision making in business organizations. *American Economic Review*, *69*(4), 493–513.

Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). Cambridge, MA: MIT Press.

Simon, H. A. (1983). *Reason in human affairs*. Stanford, CA: Stanford University Press.

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. 373–398). Hillsdale, NJ: Erlbaum.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*(1), 1–20.

Simon, H. A. (1991). *Models of my life*. New York, NY: Basic Books.

Simon, H. A. (1992). What is an "explanation" of behavior? *Psychological Science*, *3*(3), 150–161.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge, MA: MIT Press. (Original work published 1969)

Şimşek, Ö. (2013). Linear decision rule as aspiration for simple decision heuristics. *Advances in Neural Information Processing Systems*, *26*, 2904–2912.

Smith, V. L. (2003). Constructivist and ecological rationality in economics. *American Economic Review*, *93*(3), 465–508.

Spiliopoulos, L., & Hertwig, R. (2020). A map of ecologically rational heuristics for uncertain strategic worlds. *Psychological Review*, *127*(2), 245–280.

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford, England: Clarendon Press.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, *69*(3), 213–225.

Sunstein, C. R. (Ed.). (2000). *Behavioral law and economics*. Cambridge, England: Cambridge University Press.

Thaler, R. H. (2016). Behavioral economics: Past, present, and future. *American Economic Review*, *106*(7), 1577–1600.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. New York, NY: Simon & Schuster.

Todd, P. M., Gigerenzer, G., & ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. Oxford, England: Oxford University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

Vriend, N. J. (1996). Rational behavior and economic theory. *Journal of Economic Behavior & Organization*, *29*(2), 263–285.

Wheeler, G. (2019). Bounded rationality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/fall2019/entries/bounded-rationality/

This is a section of doi:10.7551/mitpress/11252.001.0001

The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.

The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455