BOUNDED RATIONALITY

GREGORY WHEELER
FRANKFURT SCHOOL OF FINANCE & MANAGEMENT
g.wheeler@fs.de

DRAFT OF September 1, 2018

Herbert Simon introduced the term 'bounded rationality' (Simon 1957, p. 198) as a shorthand for his brief against neoclassical economics and his call to replace the perfect rationality assumptions of *homo economicus* with a conception of rationality tailored to cognitively limited agents.

> 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, including man, in the kinds of environments in which such organisms exist (Simon 1955a).

'Bounded rationality' has since come to refer to a wide range of descriptive, normative, and prescriptive accounts of effective behavior which depart from the assumptions of perfect rationality. This entry aims to highlight key contributions—from the decision sciences, economics, cognitive- and neuropsychology, biology, computer science, and philosophy—to our current understanding of bounded rationality.

CONTENTS

1 Homo Economicus and Expected Utility Theory

Bounded rationality has come to broadly encompass models of effective behavior that weaken, or reject altogether, the idealized conditions of perfect rationality assumed by models of economic man. In this section we state what models of economic man are committed to and their relationship to expected utility theory. In later sections we review proposals for departing from expected utility theory.

The perfect rationality of *homo economicus* imagines a hypothetical agent who has complete information about the options available for choice, perfect foresight of the consequences from choosing those options, and the wherewithal to solve an optimization problem (typically of considerable complexity) that identifies an option which maximizes the agent's personal utility. The meaning of 'economic man' has evolved from John Stuart Mill's description of a hypothetical, self-interested individual who seeks to maximize his personal utility (1844); to Jevon's mathematization of marginal utility to model an economic consumer (1871); to Frank Knight's portrayal of the *slot-machine man* of neo-classical economics (1921), which is Jevon's *calculator man* augmented with perfect foresight and determinately specified risk; to the modern conception of an economically rational economic agent conceived in terms of Paul Samuelson's *revealed preference* formulation of utility (1947) which, together with von Neumann and Morgenstern's axiomatization (1944), changed the focus of economic modeling from reasoning behavior to choice behavior.

Modern economic theory begins with the observation that human beings like some consequences better than others, even if they only assess those consequences hypothetically. A perfectly rational person, according to the canonical paradigm of synchronic decision making under risk, is one whose comparative assessments of a set of consequences satisfies the recommendation to maximize expected utility. Yet, this recommendation to maximize expected utility presupposes that qualitative comparative judgements of those consequences (i.e., preferences) are structured in such a way (i.e., satisfy specific axioms) so as to admit a mathematical representation that places those objects of comparison on the real number line (i.e., as inequalities of mathematical expectations), ordered from worst to best. This structuring of preference through axioms to admit a numerical representation is the subject of expected utility theory.

1.1 Expected Utility Theory

We present here one such axiom system to derive expected utility theory, a simple set of axioms for the binary relation $\succeq$, which represents the relation "is weakly preferred to". The objects of comparison for this axiomatization are *prospects*, which associate probabilities to a fixed set of consequences, where both probabilities and consequences are known to the agent. To illustrate, the prospect ($-$€10,$1/2$; €20,$1/2$) concerns two consequences, *losing 10 Euros* and *winning 20 Euros*, each assigned the probability one-half. A rational agent will prefer this prospect to another with the same consequences but greater chance of losing than winning, such as ($-$€10,$2/3$; €20,$1/3$), assuming his aim is to maximize his financial welfare. More generally, suppose that $X = \{x_1, x_2, \ldots, x_n\}$ is a mutually exclusive and exhaustive set of consequences and that p_i denotes the probability of x_i, where each $p_i \geq 0$ and $\sum_i^n p_i = 1$.

A prospect P is simply the set of consequence-probability pairs, $P = (x_1, p_1;\ x_2, p_2; \ldots;\ x_n, p_n)$. By convention, a prospect's consequence-probability pairs are ordered by the value of each consequence, from least favorable to most. When prospects P, Q, R are comparable under a specific preference relation, $\succeq$, and the (ordered) set of consequences X is fixed, then prospects may be simply represented by a vector of probabilities.

The *expected utility hypothesis* (Bernoulli 1738) states that rational agents ought to maximize expected utility. If your qualitative preferences $\succeq$ over prospects satisfy the following three constraints, *ordering*, *continuity*, and *independence*, then your preferences will maximize expected utility (von Neumann and Morgenstern 1944).

A1. **Ordering**. The ordering condition states that preferences are both *complete* and *transitive*. For all prospects P, Q, completeness entails that either $P \succeq Q$, $Q \succeq P$, or both $Q \succeq P$ and $Q \succeq P$, written $P \sim Q$. For all prospects P, Q, R, transitivity entails that if $P \succeq Q$ and $Q \succeq R$, then $P \succeq R$.

A2. **Archimedean**. For all prospects P, Q, R such that $P \succeq Q$ and $Q \succeq R$, then there exists some $p \in (0,1)$ such that $(P, p;\ R, (1-p)) \sim Q$, where $(P, p; R, (1-p))$ is the *compound prospect* that yields the prospect P as a consequence with probability p or yields the prospect R with probability $1-p$.[1]

A3. **Independence**. For all prospects P, Q, R, if $P \succeq Q$, then $(P, p;\ R, (1-p)) \succeq (Q, p;\ R, (1-p))$ for all p.

Specifically, if A1, A2, and A3 hold, then there is a real-valued function $V(\cdot)$ of the form

$$V(P) = \sum_i (p_i \cdot u(x_i)) \tag{1}$$

where P is any prospect and $u(\cdot)$ is a von Neumann and Morgenstern utility function defined on the set of consequences X, such that $P \succeq Q$ if and only if $V(P) \geq V(Q)$. In other words, if your qualitative comparative judgements of prospects at a given time satisfy A1, A2, and A3, then those qualitative judgments are representable numerically by inequalities of functions of the form $V(\cdot)$, yielding a logical calculus on an interval scale for determining the consequences of your qualitative comparative judgments at that time.

1.2 Axiomatic Departures from Expected Utility Theory

It is commonplace to explore alternatives to an axiomatic system and expected utility theory is no exception. To be clear, not all departures from expected utility theory are candidates for modeling bounded rationality. Nevertheless, some confusion and misguided rhetoric over how to approach the problem of modeling bounded rationality stems from unfamiliarity with the breadth of contemporary statistical decision theory. Here we highlight some axiomatic departures from expected utility theory that are motivated by bounded rationality considerations, all framed in terms of our particular axiomatization from Section 1.1.

Alternatives to A1. Weakening the ordering axiom introduces the possibility for an agent to forgo comparing a pair of alternatives, an idea both Keynes and Knight advocated (Keynes 1921; Knight 1921). Specifically, dropping the completeness axiom allows an agent to be in a position to neither prefer one option to another nor be indifferent between the two (Koopman 1940; Aumann 1962; Fishburn 1982). Decisiveness, which the completeness axiom encodes, is more mathematical convenience than principle of rationality. The question, which is the question that every proposed axiomatic system faces, is what logically follows from a system which allows for incomplete preferences. Led by (Aumann 1962), early

[1] For compound prospects with an uncountable number of consequences, the Archimedean condition is replaced by a **continuity** condition, which maintains that $\forall p \in P$ are closed in the topology of weak convergence.

axiomatizations of rational incomplete preferences were suggested by (Giles 1976) and (Giron and Rios 1980), and later studied by (Karni 1985), (Bewley 2002), (Walley 1991), (Seidenfeld, Schervish, and Kadane 1995), (Ok 2002), (Nau 2006), (Galaabaatar and Karni 2013) and (Zaffalon and Miranda 2015). In addition to accommodating indecision, such systems also allow for you to reason about someone else's (possibly) complete preferences when your information about that other agent's preferences is incomplete.

Dropping transitivity limits extendability of elicited preferences (Luce and Raiffa 1957), since the omission of transitivity as an axiomatic constraint allows for cycles and preference reversals. Although violations of transitivity have been long considered both commonplace and a sign of human irrationality (May 1954; Tversky 1969), reassessments of the experimental evidence challenge this received view (Mongin 2000; Regenwetter, Dana, and Davis-Stober 2011). The axioms impose synchronic consistency constraints on preferences, whereas the experimental evidence for violations of transitivity commonly conflate dynamic and synchronic consistency (Regenwetter, Dana, and Davis-Stober 2011). Specifically, a person's preferences at one moment in time that are inconsistent with his preferences at another time is no evidence for that person holding logically inconsistent preferences at a single moment in time. Arguments to limit the scope of transitivity in normative accounts of rational preference similarly point to diachronic or group preferences, which likewise do not contradict the axioms (Kyburg 1978; Schick 1986; Anand 1987; Bar Hillel and Margalit 1988). Arguments that point to psychological processes or algorithms that admit cycles or reversals of preference over time also point to a misapplication of, rather than a counter-example to, the ordering condition. Finally, for decisions that involve explicit comparisons of options over time, violating transitivity may be rational. For example, given the goal of maximizing the rate of food gain, an organism's current food options may reveal information about food availability in the near future by indicating that a current option may soon disappear or that a better option may soon reappear. Information about availability of options over time can, and sometimes does, warrant non-transitive choice behavior over time that nevertheless maximizes food gain (McNamara, Trimmer, and Houston 2014).

Alternatives to A2. Dropping the Archimedean axiom allows for an agent to have *lexicographic preferences* (Blume, Brandenburger, and Dekel 1991); that is, the omission of A2 allows the possibility for an agent to prefer one option infinitely more than another. One motivation for developing a non-Archemedean version of expected utility theory is to address a gap in the foundations of the standard subjective utility framework that prevents a full reconciliation of *admissibility* (i.e., the principle that one ought not select a weakly dominated option for choice) with *full conditional preferences* (i.e., that for any event, there is a well-defined conditional probability to represent the agent's conditional preferences) (Pedersen 2014). Specifically, the standard subjective expected utility account cannot accommodate conditioning on zero-probability events, which is of particular importance to game theory (Hammond 1994). Non-Archimedean variants of expected utility theory turn to techniques from nonstandard analysis (Goldblatt 1998), full conditional probabilities (Renyi 1955; Popper 1959; Dubins 1975; Coletti and Scozzafava 2002), and lexicographic probabilities (Halpern 2010; Brickhill and Horsten 2016), and are all linked to imprecise probability theory (Wheeler and Cozman 2018).

Non-compensatory single-cue decision models, such as the Take-the-Best heuristic (Section 7.2), appeal to lexicographically ordered cues, and admit a numerical representation in terms of non-Archimedean expectations (Arló-Costa and Pedersen 2011).

Alternatives to A3. A1 and A2 together entail that $V(\cdot)$ assigns a real-valued index to prospects such that $P \succeq Q$ if and only if $V(P) \geq V(Q)$. The independence axiom, A3, encodes a separability property for choice, one that ensures that expected utilities are linear in probabilities. Motivations for dropping the independence axiom stem from difficulties in applying expected utility theory to describe choice behavior, including an early observation that humans evaluate possible losses and possible gains differently. Although expected utility theory can represent a person who either gambles or purchases insurance, Friedman and Savage remarked in their early critique von Neumann and Morgenstern's axiomization, it

cannot simultaneously do both (Friedman and Savage 1948).

The principle of loss aversion (Kahneman and Tversky 1979; Rabin 2000) suggests that the subjective weight that we assign to potential losses is larger than those we assign to potential gains. For example, the *endowment effect* (Thaler 1980)—the observation that people tend to view the value of a good higher when viewed as a potential loss than when viewed as a potential gain—is supported by neurological evidence for gains and losses being processed by different regions of the brain (Rick 2011). However, even granting the affective differences in how we process losses and gains, those differences do not necessarily translate to a general "negativity bias" (Baumeister, Bratslavsky, and Finkenauer 2001) in choice behavior (Hochman and Yechiam 2011; Yechiam and Hochman 2014). Yechiam and colleagues report experiments in which participants do not exhibit loss aversion in their choices, such as cases in which participants respond to repetitive situations that issue losses and gains and single-case decisions involving small stakes. That said, observations of risk aversion (Allais 1953) and ambiguity aversion (Ellsberg 1961) have led to alternatives to expected utility theory, all of which abandon A3. Those alternative approaches include prospect theory (Section 2.4), regret theory (Bell 1982; Loomes and Sugden 1982), and rank-dependent expected utility (Quiggin 1982).

Most models of bounded rationality do not even fit into this broad axiomatic family just outlined. One reason is that bounded rationality has historically emphasized the procedures, algorithms, or psychological processes involved in making a decision, rendering a judgment, or securing a goal (Section 2). Samuelson's shift from reasoning behavior to choice behavior abstracted away precisely these details, however, treating them as outside the scope of rational choice theory. For Simon, that was precisely the problem. A second reason is that bounded rationality often focuses on adaptive behavior suited to an organism's environment (Section 3). Since ecological modeling involves goal-directed behavior mitigated by the constitution of the organism and stable features of its environment, focusing on (synchronically) coherent comparative judgments is often not, directly at least, the best way to frame the problem.

That said, one should be cautious about generalizations sometimes made about the limited role of decision theoretic tools in the study of bounded rationality. Decision theory—broadly construed to include statistical decision theory (Berger 1980)—offers a powerful mathematical toolbox even though historically, particularly in its canonical form, it has traded in psychological myths such as "degrees of belief" and logical omniscience (Section 1.3). One benefit of studying axiomatic departures from expected utility theory is to loosen the grip of Bayesian dogma to expand the range of possibilities for applying a growing body of practical and powerful mathematical methods.

1.3 Limits to Logical Omniscience

Most formal models of judgment and decision making entail *logical omniscience*—complete knowledge of all that logically follows from one's current commitments combined with any set of options considered for choice—which is as psychologically unrealistic as it is difficult, technically, to avoid (Stalnaker 1991). A descriptive theory that presumes or a prescriptive theory that recommends to disbelieve a claim when the evidence is logically inconsistent, for example, will be unworkable when the belief in question is sufficiently complicated for all but logically omniscient agents, even for non-omniscient agents that nevertheless have access to unlimited computational resources (Kelly and Schulte 1995).

The problem of logical omniscience is particularly acute for expected utility theory in general, and the theory of subjective probability in particular. For the postulates of subjective probability imply that an agent knows all the logical consequences of her commitments, thereby mandating logical omniscience. This limits the applicability of the theory, however. For example, it prohibits having uncertain judgments about mathematical and logical statements. In an article from 1967, "Difficulties in the theory of personal probability," reported in (Hacking 1967) and (Seidenfeld, Schervish, and Kadane 2012) but misprinted in (Savage 1967), Savage raises the problem of logical omniscience for the subjective theory of probability:

> The analysis should be careful not to prove too much; for some departures from theory are inevitable, and some even laudable. For example, a person required to risk money on

> a remote digit of π would, in order to comply fully with the theory, have to compute that digit, though this would really be wasteful if the cost of computation were more than the prize involved. For the postulates of the theory imply that you should behave in accordance with the logical implication of all that you know. Is it possible to improve the theory in this respect, making allowances within it for the cost of thinking, or would that entail paradox, as I am inclined to believe but unable to demonstrate? (Savage 1967, excerpted from Savage's prepublished draft. See notes in Seidenfeld et al., 2012)

Responses to Savage's problem include a game-theoretic treatment proposed by I.J. Good (1983), which swaps the extensional variable that is necessarily true for an intensional variable representing an accomplice who knows the necessary truth but withholds enough information from you for you to be (coherently) uncertain about what he knows. This trick changes the subject of your uncertainty, from a necessarily true proposition that you cannot coherently doubt to a coherent guessing game about that truth facilitated by your accomplice's incomplete description. Another response sticks to the classical line that failures of logical omniscience are deviations from the normative standard of perfect rationality but introduces an index for incoherence to accommodate reasoning with incoherent probability assessments (Schervish, Seidenfeld, and Kadane 2012). A third approach, suggested by de Finetti (1974), is to restrict possible states of affairs to observable states with a finite verifiable procedure—which may rule out theoretical states or any other that does not admit a verification protocol. Originally, what de Finetti was after was a principled way to construct a partition over possible outcomes to distinguish serious possible outcomes of an experiment from wildly implausible but logically possible outcomes, yielding a method for distinguishing between genuine doubt and mere "paper doubts" (Peirce 1955). Other proposals follow de Finetti's line by tightening the admissibility criteria and include *epistemically possible* events, which are events that are logically consistent with the agent's available information; *apparently possible* events, which include any event by default unless the agent has determined that it is inconsistent with his information; and *pragmatically possible* events, which only includes events that are judged sufficiently important (Walley 1991, §2.1).

The notion of *apparently possible* refers to a procedure for determining inconsistency, which is a form of bounded procedural rationality (Section 2). The challenges of avoiding paradox, which Savage alludes to, are formidable. However, work on bounded fragments of Peano arithmetic (Parikh 1971) provided coherent foundations for exploring these ideas, which has been taken up specifically to formulate bounded-extensions of default logic for *apparent possibility* (Wheeler 2004) and more generally in models of *computational rationality* (Lewis, Howes, and Singh 2014).

1.4 Descriptions, Prescriptions, and Normative Standards

It is commonplace to contrast how people render judgments, or make decisions, from how they ought to do so. However, interest in cognitive processes, mechanisms, and algorithms of boundedly rational judgment and decision making suggests that we instead distinguish among three aims of inquiry rather than these two. Briefly, a *descriptive theory* aims to explain or predict what judgments or decisions people in fact make; a *prescriptive theory* aims to explain or recommend what judgments or decisions people ought to make; a *normative theory* aims to specify a normative standard to use in evaluating the effectiveness of a judgement or decision.

To illustrate each type, consider a domain where differences between these three lines of inquiry are especially clear: arithmetic. A descriptive theory of arithmetic might concern the psychology of arithmetical reasoning, a model of approximate numeracy in animals, or an algorithm for implementing arbitrary-precision arithmetic on a digital computer. The normative standard of full arithmetic is Peano's axiomatization of arithmetic, which distills natural number arithmetic down to a function for one number succeeding another and mathematical induction. But one might also consider Robinson's induction-free fragment of Peano arithmetic (Tarski, Mostowski, and Robinson 1953) or axioms for some system of cardinal arithmetic in the hierarchy for large cardinals. A prescriptive theory for arithmetic will reference both a fixed normative standard and relevant facts about the arithmetical capabilities of the organism or

machine performing arithmetic. A curriculum for improving the arithmetical performance of elementary school children will differ from one designed to improve the performance of adults. Even though the normative standard of Peano arithmetic is the same for both children and adults, stable psychological differences in these two populations may warrant prescribing different approaches for improving their arithmetic. Continuing, even though Peano's axioms are the normative standard for full arithmetic, nobody would prescribe Peano's axioms for the purpose of improving anyone's sums. There is no mistaking Peano's axioms for a descriptive theory of arithmetical reasoning, either. Even so, a descriptive theory of arithmetic will presuppose the Peano axioms as the normative standard for full arithmetic, even if only implicitly. In describing how people sum two numbers, after all, one presumes that they are attempting to sum two numbers rather than concatenate them, count out in sequence, or send a message in code.

Finally, imagine an effective pedagogy for teaching arithmetic to children is known and we wish to introduce children to cardinal arithmetic. A reasonable start on a prescriptive theory for cardinal arithmetic for children might be to adapt as much of the successful pedagogy for full arithmetic as possible while anticipating that some of those methods will not survive the change in normative standards from Peano to (say) ZFC+. Some of those differences can be seen as a direct consequence of the change from one standard to another, while other differences may arise unexpectedly from the observed interplay between the change in task, that is, from performing full arithmetic to performing cardinal arithmetic, and the psychological capabilities of children to perform each task.

To be sure, there are important differences between arithmetic and rational behavior. The objects of arithmetic, numerals and the numbers they refer to, are relatively clear cut, whereas the objects of rational behavior vary even when the same theoretical machinery is used. Return to expected utility theory as an example. An agent may be viewed as deliberating over options with the aim to choose one that maximizes his personal welfare, or viewed to act as if he deliberately does so without actually doing so, or understood to do nothing of the kind but to instead be a bit part player in the population fitness of his kind.

Separating the question of how to choose a normative standard from questions about how to evaluate or describe behavior is an important tool to reduce misunderstandings that arise in discussions of bounded rationality. Even though Peano's axioms would never be prescribed to improve, nor proposed to describe, arithmetical reasoning, it does not follow that the Peano axioms of arithmetic are irrelevant to descriptive and prescriptive theories of arithmetic. While it remains an open question whether the normative standards for human rational behavior admit axiomatization, there should be little doubt over the positive role that clear normative standards play in advancing our understanding of how people render judgments, or make decisions, and how they ought to do so.

2 The Emergence of Procedural Rationality

Simon thought the shift in focus from reasoning behavior to choice behavior was a mistake. Since,in the 1950s, little was known about the processes involved in making judgments or reaching decisions, we were not in the position to freely abstract away all of those features from our mathematical models. Yet, this ignorance also raised the question of how to proceed. The answer was to attend to the costs in effort involved operating a procedure for making decisions and comparing those costs to the resources available to the organism using the procedure and, conversely, to compare how well an organism performs in terms of accuracy (Section 8.2) with its limited cognitive resources in order to investigate models with comparable levels of accuracy within those resource bounds. Effectively managing the trade-off between the costs and quality of a decision involves another type of rationality, which Simon later called *procedural rationality* (Simon 1976, p. 69).

In this section we highlight early, key contributions to modeling procedures for boundedly rational judgment and decision-making, including the origins of the *accuracy-effort trade-off*, Simon's *satisficing* strategy, *improper linear models*, and the earliest effort to systematize several features of high-level, cognitive judgment and decision-making, *cumulative prospect theory*.

2.1 Accuracy and Effort

Herbert Simon and I.J. Good were each among the first to call attention to the cognitive demands of subjective expected utility theory, although neither one in his early writings abandoned the principle of expected utility as the normative standard for rational choice. Good, for instance, referred to the recommendation to maximize expected utility as the *ordinary principle* of rationality, whereas Simon called the principle *objective rationality* and considered it the central tenant of *global rationality*. The rules of rational behavior are costly to operate in both time and effort, Good observed, so real agents have an interest in minimizing those costs (Good 1952, §7(i)). Efficiency dictates that one choose from available alternatives an option that yields the largest result given the resources available, which Simon emphasized is not necessarily an option that yields the largest result overall (Simon 1947, p. 79). So reasoning judged deficient without considering the associated costs may be found meritorious once all those costs are accounted for—a conclusion that a range of authors soon came to endorse, including Amos Tversky:

> It seems impossible to reach any definitive conclusions concerning human rationality in the absence of a detailed analysis of the sensitivity of the criterion and the cost involved in evaluating the alternatives. When the difficulty (or the costs) of the evaluations and the consistency (or the error) of the judgments are taken into account, a [transitivity-violating method] may prove superior (Tversky 1969).

Balancing the quality of a decision against its costs soon became a popular conception of bounded rationality, particularly in economics (Stigler 1961), where it remains commonplace to formulate boundedly rational decision-making as a constrained optimization problem. On this view boundedly rational agents are utility maximizers after all, once all the constraints are made clear (Arrow 2004). Another reason for the popularity of this conception of bounded rationality is its compatibility with Milton Friedman's *as if* methodology (Friedman 1953), which licenses models of behavior that ignore the causal factors underpinning judgment and decision making. To say that an agent behaves *as if* he is a utility maximizer is at once to concede that he is not but that his behavior proceeds as if he were. Similarly, to say that an agent behaves as if he is a utility maximizer under certain constraints is to concede that he does not solve constrained optimization problems but nevertheless behaves as if he did so.

Simon's focus on computationally efficient methods that yield solutions that are good enough contrasts with Friedman's as if methodology, since evaluating whether a solution is "good enough", in Simon's terms, involves search procedures, stopping criteria, and how information is integrated in the course of making a decision. Simon offers several examples to motivate inquiry into computationally efficient methods. Here is one. Applying the game-theoretic minimax algorithm to the game of chess calls for evaluating more chess positions than the number of molecules in the universe (Simon 1957, p. 6). Yet if the game of chess is beyond the reach of exact computation, why should we expect everyday problems to be any more tractable? Simon's question is to explain how human beings manage to solve complicated problems in an uncertain world given their meager resources. Answering Simon's question, as opposed to applying Friedman's method to fit a constrained optimization model to observed behavior, is to demand a model with better predictive power concerning boundedly rational judgment and decision making. In pressing this question of how human beings solve uncertain inference problems, Simon opened two lines of inquiry that continue to today, namely:

1. How do human beings actually make decisions "in the wild"?
2. How can the standard theories of global rationality be simplified to render them more tractable?

Simon's earliest efforts aimed to answer the second question with, owing to the dearth of psychological knowledge at the time about how people actually make decisions, only a layman's "acquaintance with the gross characteristics of human choice" (Simon 1955a, p. 100). His proposal was to replace the optimization problem of maximizing expected utility with a simpler decision criterion he called *satisficing*, and by models with better predictive power more generally.

2.2 Satisficing

Satisficing is the strategy of considering the options available to you for choice until you find one that meets or exceeds a predefined threshold—your aspiration level—for a minimally acceptable outcome. Although Simon originally thought of procedural rationality as a poor approximation of global rationality, and thus viewed the study of bounded rationality to concern "the behavior of human beings who *satisfice* because they have not the wits to *maximize*" (Simon 1957, p. xxiv), there are a range of applications of satisficing models to sequential choice problems, aggregation problems, and high-dimensional optimization problems, which are increasingly common in machine learning.

Given a specification of what will count as a good-enough outcome, satisficing replaces the optimization objective from expected utility theory of selecting an undominated outcome with the objective of picking an option that meets your aspirations. The model has since been applied to business (Bazerman and Moore 2008; Puranam, Stieglitz, Osman, and Pillutla 2015), mate selection (Todd and Miller 1999) and other practical sequential-choice problems, like selecting a parking spot (Hutchinson, Fanselow, and Todd 2012). Ignoring the procedural aspects of Simon's original formulation of satisficing, if one has a fixed aspirational level for a given decision problem, then admissible choices from satisficing can be captured by so-called ε-efficiency methods (Loridan 1984; White 1986).

Hybrid optimization-satisficing techniques are used in machine learning when many metrics are available but no sound or practical method is available for combining them into a single value. Instead, hybrid optimization-satisficing methods select one metric to optimize and satisfice the remainder. For example, a machine learning classifier might optimize accuracy (i.e., maximize the proportion of examples for which the model yields the correct output; see Section 8.2) but set aspiration levels for the false positive rate, coverage, and runtime.

Selten's *aspiration adaption theory* models decision tasks as problems with multiple incomparable goals that resist aggregation into a complete preference order over all alternatives (Selten 1998). Instead, the decision-maker will have a vector of goal variables, where those vectors are comparable by weak dominance. If vector A and vector B are possible assignments for my goals, then A dominates vector B if there is no goal in the sequence in which B assigns a value that is strictly less than A, and there is some goal for which A assigns a value strictly greater than B. Selten's model imagines an aspiration level for each goal, which itself can be adjusted upward or downwards depending on the set of feasible (admissible) options. Aspiration adaption theory is a highly procedural and local account in the tradition of Newell and Simon's approach to human problem solving (Newell and Simon 1972), although it was not initially offered as a psychological process model. Analogous approaches have been explored in the AI planning literature (Bonet and Geffner 2001; Ghallab, Nau, and Traverso 2016).

2.3 Proper and Improper Linear Models

Proper linear models represent another important class of optimization models. A proper linear model is one where predictor variables are assigned weights, which are selected so that the linear combination of those weighted predictor variables optimally predicts a target variable of interest. For example, linear regression is a proper linear model that selects weights such that the squared "distance" between the model's predicted value of the target variable and the actual value (given in the data set) is minimized.

Paul Meehl's review in the 1950s of psychological studies using statistical methods versus clinical judgment cemented the statistical turn in psychology (Meehl 1954). Meehl's review found that studies involving the prediction of a numerical target variable from numerical predictors is better done by a proper linear model than by the intuitive judgment of clinicians. Concurrently, the psychologist Kenneth Hammond formulated Brunswik's lens model (Section 3.2) as a composition of proper linear models to model the differences between clinical versus statistical predictions (Hammond 1955). Proper linear models have since become a workhorse in cognitive psychology in areas that include decision analysis (Keeney and Raiffa 1976; Kaufmann and Wittmann 2016), causal inference (Waldmann, Holyoak, and Fratianne 1995; Spirtes 2010), and response-times to choice (Brown and Heathcote 2008; Turner, Rodriguez, Norcia, McClure, and Steyvers 2016).

Robin Dawes, returning to Meehl's question about statistical versus clinical predictions, found that even improper linear models perform better than clinical intuition (Dawes 1979). The distinguishing feature of improper linear models is that the weights of a linear model are selected by some non-optimal method. For instance, equal weights might be assigned to the predictor variables to afford each equal weight or unit-weights assigned, such as -1 or 1, to tally features supporting a positive or negative prediction, respectively. As an example, Dawes proposed an improper model to predict subjective ratings of marital happiness by couples based on the difference between their rates of lovemaking and fighting. The results? Among the thirty happily married couples, two argued more than they had intercourse. Yet all twelve unhappy couples fought more frequently. And those results replicated in other laboratories studying human sexuality in the 1970s. Both equal-weight regression and unit-weight *tallying* have since been found to commonly outperform proper linear models on small data sets. Although no simple improper linear model performs well across all common benchmark datasets, for almost every data set in the benchmark there is some simple improper model that performs well in predictive accuracy (Lichtenberg and Özgür Simsek 2016). This observation, and many others in the heuristics literature, points to biases of simplified models that can lead to better predictions when used in the right circumstances (Section 4).

Dawes's original point was not that improper linear models outperform proper linear models in terms of accuracy, but rather that they are more efficient and (often) close approximations of proper linear models. "The statistical model may integrate the information in an optimal manner," Dawes observed, "but it is always the individual ... who chooses variables" (Dawes 1979, p. 573). Moreover, Dawes argued that it takes human judgment to know the direction of influence between predictor variables and target variables, which includes the knowledge of how to numerically code those variables to make this direction clear. Recent advances in machine learning chip away at Dawes's claims about the unique role of human judgment, and results from Gigerenzer's ABC Group about unit-weight tallying outperforming linear regression in out-of-sample prediction tasks with small samples is an instance of improper linear models outperforming proper linear models (Czerlinski, Gigerenzer, and Goldstein 1999). Nevertheless, Dawes's general observation about the relative importance of variable selection over variable weighting stands (Katsikopoulos, Schooler, and Hertwig 2010).

2.4 Cumulative Prospect Theory

If both satisficing and improper linear models are examples addressing Simon's second question at the start of this section—namely, how to simplify existing models to render them both tractable and effective—then Daniel Kahneman and Amos Tversky's *cumulative prospect theory* is among the first models to directly incorporate knowledge about how humans actually make decisions.

In our discussion in Section 1.1 about alternatives to the Independence Axiom, (A3), we mentioned several observed features of human choice behavior that stand at odds with the prescriptions of expected utility theory. Kahneman and Tversky developed prospect theory around four of those observations about human decision-making (Kahneman and Tversky 1979; Wakker 2010).

1. **Reference Dependence**. Rather than make decisions by comparing the absolute magnitudes of welfare, as prescribed by expected utility theory, people instead tend to value prospects by their change in welfare with respect to a reference point. This reference point can be a person's current state of wealth, an aspiration level, or a hypothetical point of reference from which to evaluate options. The intuition behind reference dependence is that our sensory organs have evolved to detect changes in sensory stimuli rather than store and compare absolute values of stimuli. Therefore, the argument goes, we should expect to see the cognitive mechanisms involved in decision-making to inherit this sensitivity to changes in perceptual attributes values.

   In prospect theory, reference dependence is reflected by utility changing sign at the origin of the valuation curve $v(\cdot)$ in Figure 1(a). The x-axis represents gains (right side) and losses (left side) in euros, and y-axis plots the value placed on relative gains and losses by a valuation function $v(\cdot)$, which is fit to experimental data on people's choice behavior.

2. **Loss Aversion**. People are more sensitive to losses than gains of the same magnitude; the thrill of victory does not measure up to the agony of defeat. So, Kahneman and Tversky maintained, people will prefer an option that does not incure a loss to an alternative option that yields an equivalent gain. The disparity in how potential gains and losses are evaluated also accounts for the endowment effect, which is the tendency for people to value a good that they own more than a comparatively valued substitute (Thaler 1980).

   In prospect theory, loss aversion appears in Figure 1(a) in the (roughly) steeper slope of $v(\cdot)$ to the left of the origin, representing losses relative to the subject's reference point, than the slope of $v(\cdot)$ for gains on the right side of the reference point. Thus, for the same magnitude of change in reward x from the reference point, there magnitude of the consequence of gaining x is less than the magnitude of losing x.

   Note that differences in affective attitudes toward, and the neurological processes responsible for processing, losses and gains do not necessarily translate to differences in people's choice behavior (Yechiam and Hochman 2014). The role and scope that loss aversion plays in judgment and decision making is less clear than was initially assumed (Section 1.2).

3. **Diminishing Returns for both Gains and Losses**. Given a fixed reference point, people's sensitivity to changes in asset values (x in Figure 1a) diminish the further one moves from that reference point, both in the domain of losses and the domain of gains. This is inconsistent with expected utility theory, even when the theory is modified to accommodate diminishing marginal utility (Friedman and Savage 1948).

   In prospect theory, the valuation function $v(\cdot)$ is concave for gains and convex for losses, representing a diminishing sensitivity to both gains and losses. Expected utility theory can be made to accommodate sensitivity effects, but the utility function is typically either strictly concave or strictly convex, not both.

4. **Probability Weighting**. Finally, for known exogenous probabilities, people do not calibrate their subjective probabilities by direct inference (Levi 1977), but instead systematically underweight high-probability events and overweight low-probability events, with a cross-over point of approximately one-third (Figure 1b). Thus, changes in very small or very large probabilities have greater impact on the evaluation of prospects than they would under expected utility theory. People are willing to pay more to reduce the number of bullets in the chamber of a gun from 1 to 0 than from 4 bullets to 3 in a hypothetical game of Russian roulette.

   Figure 1(b) plots the median values for the probability weighting function $w(\cdot)$ that takes the exogenous probability p associated with prospects, as reported in (Tversky and Kahneman 1992). Roughly, below probability values of one-third people overestimate the probability of an outcome (consequence), and above probability one-third people tend to underestimate the probability of an outcome occurring. Traditionally, overweighting is thought to concern the systematic miscalibration of people's subjective estimates of outcomes against a known exogenous probability, p, serving as the reference standard. In support of this view, miscalibration appears to disappear when people learn a distribution through sampling instead of learning identical statistics by description (Hertwig, Barron, Weber, and Erev 2004). Miscalibration in this context ought to be distinguished from overestimating or underestimating subjective probabilities when the relevant statistics are not supplied as part of the decision task. For example, televised images of the aftermath of airplane crashes lead to an overestimation of the low-probability event of commercial airplanes crashing. Even though a person's subjective probability of the risk of a commercial airline crash would be too high given the statistics, the mechanism responsible is different: here the *recency* or *availability* of images from the evening news is to blame for scaring him out of his wits, not the sober fumbling of a statistics table. An alternative view maintains that people understand that their weighted probabilities are different than the exogenous probability but nevertheless prefer to act as if the exogenous probability were so weighted (Wakker 2010). On this

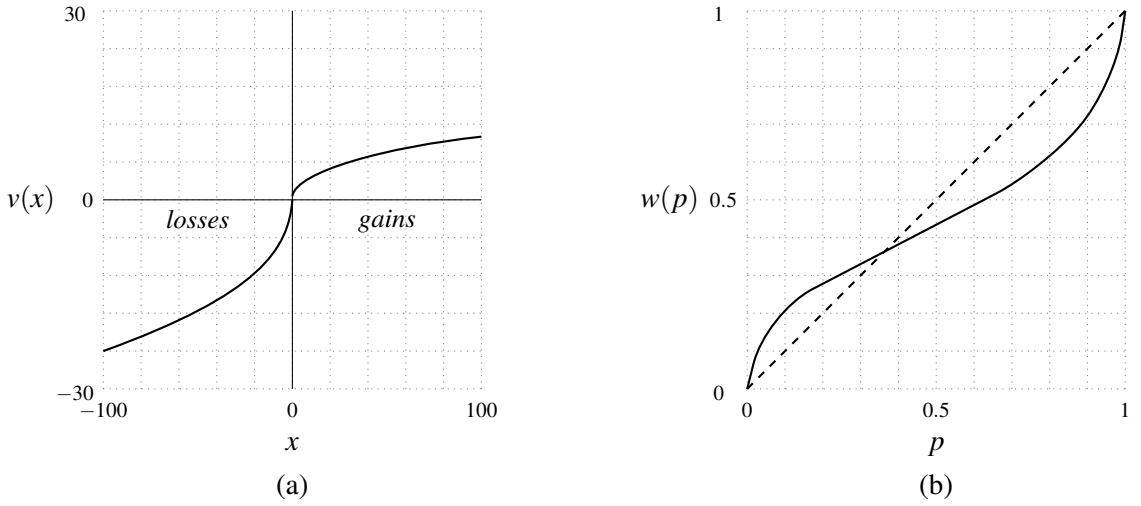


Figure 1: (a) plots the value function $v(\cdot)$ applied to consequences of a prospect; (b) plots the median value of the probability weighting function $w(\cdot)$ applied to positive prospects of the form $(x, p; 0, 1-p)$ with probability p.

view, probability weighting is not a (mistaken) belief but a preference.

Prospect theory incorporates these components into models of human choice under risk by first identifying a reference point that either refers to the status quo or some other aspiration level. The consequences of the options under consideration then are framed in terms of deviations from this reference point. Extreme probabilities are simplified by rounding off, which yields miscalibration of the given, exogenous probabilities. Dominance reasoning is then applied, where dominated alternatives are eliminated from choice, along with additional steps to separate options without risk, probabilities associated with a specific outcome are combined, and a version of eliminating irrelevant alternatives is applied (Kahneman and Tversky 1979, pp. 284–285).

Nevertheless, the prospect theory comes with problems. For example, a shift of probability from less favorable outcomes to more favorable outcomes ought to yield a better prospect, all things considered, but the original prospect theory violates this principle of stochastic dominance. *Cumulative prospect theory* satisfies stochastic dominance, however, by appealing to a rank-dependent method for transforming probabilities (Quiggin 1982). For a review of the differences between prospect theory and cumulative prospect theory, along with an axiomatization of cumulative prospect theory, see (Fennema and Wakker 1997).

3 The Emergence of Ecological Rationality

Imagine a meadow whose plants are loaded with insects and few are in flight. Then, this meadow is a more favorable environment for a bird that gleans rather than hawks. In a similar fashion, a decision-making environment might be more favorable for one decision-making strategy than for another. Just as it would be "irrational" for a bird to hawk rather than glean, given the choice in this meadow, so too what may be an irrational decision strategy in one environment may be entirely rational in another.

If procedural rationality attaches a cost to the making of a decision, then ecological rationality locates that procedure in the world. The questions ecological rationality ask is what features of an environment can help or hinder decision making and how should we model judgment or decision-making ecologies. For example, people make causal inferences about patterns of covariation they observe—especially children, who then perform experiments testing their causal hypotheses (Glymour 2001). Unsurprisingly, people who draw the correct inferences about the true causal model do better than those who infer the wrong causal model (Meder, Mayrhofer, and Waldmann 2014). More surprising, Meder and his colleagues found that those making correct causal judgments do better than subjects who make no causal

judgments at all. And perhaps most surprising of all is that those with true causal knowledge also beat the benchmark standards in the literature which ignore causal structure entirely; the benchmarks encode, spuriously, the assumption that the best we can do is to make no causal judgments at all.

In this section and the next we will cover five important contributions to the emergence of ecological rationality. In this section, after reviewing Simon's proposal for distinguishing between *behavioral constraints* and *environmental structure*, we turn to three historically important contributions: *the lens model*, *rational analysis*, and *cultural adaptation*. Finally, in Section 4, we review the *bias-variance decomposition*, which has figured in the Fast and Frugal Heuristics literature (Section 7.2).

3.1 Behavioral Constraints and Environmental Structure

Simon thought that both behavioral constraints and environmental structure ought to figure in a theory of bounded rationality, yet he cautioned against identifying behavioral and environmental properties with features of an organism and features of its physical environment, respectively:

> we must be prepared to accept the possibility that what we call "the environment" may lie, in part, within the skin of the biological organisms. That is, some of the constraints that must be taken as givens in an optimization problem may be physiological and psychological limitations of the organism (biologically defined) itself. For example, the maximum speed at which an organism can move establishes a boundary on the set of its available behavior alternatives. Similarly, limits on computational capacity may be important constraints entering into the definition of rational choice under particular circumstances. (Simon 1955a, p. 101)

That said, what is classified as a behavioral constraint rather than an environmental affordance varies across disciplines and the theoretical tools pressed into service. For example, one computational approach to bounded rationality, *computational rationality theory* (Lewis, Howes, and Singh 2014), classifies the cost to an organism of executing an optimal program as a behavioral constraint, classifies limits on memory as an environmental constraint, and treats the costs associated with searching for an optimal program to execute as exogenous. Anderson and Schooler's study and computational modeling of human memory (Anderson and Schooler 1991) within the ACT-R framework, on the other hand, views the limits on memory and search-costs as behavioral constraints which are adaptive responses to the structure of the environment. Still another broad class of computational approaches are found in *statistical signal processing*, such as adaptive filters (Haykin 2013), which are commonplace in engineering and vision (Marr 1982; Ballard and Brown 1982). Signal processing methods typically presume the sharp distinction between device and world that Simon cautioned against, however. Still others have challenged the distinction between behavioral constraints and environmental structure by arguing that there is no clear way to separate organisms from the environments they inhabit (Gibson 1979), or by arguing that features of cognition which appear body-bound may not be necessarily so (Clark and Chalmers 1998).

Bearing in mind the different ways the distinction between behavior and environment have been drawn, and challenges to what precisely follows from drawing such a distinction, ecological approaches to rationality all endorse the thesis that the ways in which an organism manages structural features of its environment are essential to understanding how deliberation occurs and effective behavior arises. In doing so theories of bounded rationality have traditionally focused on at least some of the following features, under this rough classification:

- **Behavioral Constraints** – may refer to bounds on computation, such as the *cost of searching* the best algorithm to run, an appropriate rule to apply, or a satisficing option to choose; the *cost of executing* an optimal algorithm, appropriate rule, or satisficing choice; and *costs of storing* the data structure of an algorithm, the constitutive elements of a rule, or the objects of a decision problem.

- **Ecological Structure** – may refer to *statistical*, *topological*, or other *perceptible invariances* of the task environment that an organism is adapted to; or to *architectural features* or *biological features* of the computational processes or cognitive mechanisms responsible for effective behavior, respectively.

3.2 Brunswik's Lens Model

Egon Brunswik was among the first to apply probability and statistics to the study of human perception, and was ahead of his time in emphasizing the role ecology plays in the generalizability of psychological findings. Brunswik thought psychology ought to aim for statistical descriptions of adaptive behavior (Brunswik 1943). Instead of isolating a small number of independent variables to manipulate *systematically* to observe the effects on a dependent variable, psychological experiments ought instead to assess how an organism adapts to its environment. So, not only should experimental subjects be representative of the population, as one would presume, but the experimental situations they are subjected to ought to be representative of the environment that the subjects inhabit (Brunswik 1955). Thus, Brunswik maintained, psychological experiments ought to employ a *representative design* to preserve the causal structure of an organism's natural environment. For a review of the development of representative design and its use in the study of judgment and decision-making, see (Dhami, Hertwig, and Hoffrage 2004).

Brunswik's *lens model* is formulated around his ideas about how behavioral and environmental conditions bear on organisms perceiving proximal cues to draw inferences about some distal feature of its "natural-cultural habitat" (Brunswik 1955, p.198). To illustrate, an organism may detect the color markings (distal object) of a potential mate through contrasts in light frequencies reflecting across its retina (proximal cues). Some proximal cues will be more informative about the distal objects of interest than others, which Brunswik understood as a difference in the "objective" correlations between proximal cues and the target distal objective. The *ecological validity* of proximal cues thus refers to their capacity for providing the organism useful information about some distal object within a particular environment. Assessments of performance for an organism then amount to a comparison of the organism's actual use of cue information to the cue's information capacity.

Kenneth Hammond and colleagues (Hammond, Hursch, and Todd 1964) formulated Brunswik's lens model as a system of linear bivariate correlations, as depicted in Figure 2 (Hogarth and Karelaia 2007). Informally, Figure 2 says that the accuracy of a subject's judgment (response), Y_s, about a numerical target criterion, Y_e, given some informative cues (features) $X_1, \ldots, X_n$, is determined by the correlation between the subject's response and the target. More specifically, the linear lens model imagines two large linear systems, one for the environment, e, and another for the subject, s, which both share a set of cues, $X_1, \ldots, X_n$. Note that cues may be associated with one another, i.e., it is possible that $\rho(X_i, X_j) \neq 0$ for indices $i \neq j$ from 1 to n.

The accuracy of the subject's judgment Y_s about the target criterion value Y_e is measured by an *achievement index*, r_a, which is computed by Pearson's correlation coefficient ρ of Y_e and Y_s. The subject's predicted response $\hat{Y}_s$ to the cues is determined by the weights β_{s_i} the subject assigns to each cue X_i, and the linearity of the subject's response, R_s, measures the noise in the system, ε_s. Thus, the subject's response is conceived to be a weighted linear sum of subject-weighted cues plus noise. The analogue of *response linearity* in the environment is *environmental predictability*, R_e. The environment, on this model, is thought to be probabilistic—or "chancy" as some say. Finally, the environment-weighted sum of cues, $\hat{Y}_e$, is compared to the subject-weighted sum of cues, $\hat{Y}_s$, by a *matching index*, G.

In light of this formulation of the lens model, return to Simon's remarks concerning the classification of environmental affordance versus behavioral constraint. The conception of the lens model as a linear model is indebted to signal detection theory, which was developed to improve the accuracy of early radar systems. Thus, the model inherits from engineering a clean division between subject and environment. However, suppose for a moment that both the environmental mechanism producing the criterion value and the subject's predicted response are linear. Now consider the error-term, ε_s. That term may refer to biological constraints that are responses to adaptive pressures on the whole organism. If so, ought ε_s be classified as an environmental constraint rather than a behavioral constraint? The answer will depend

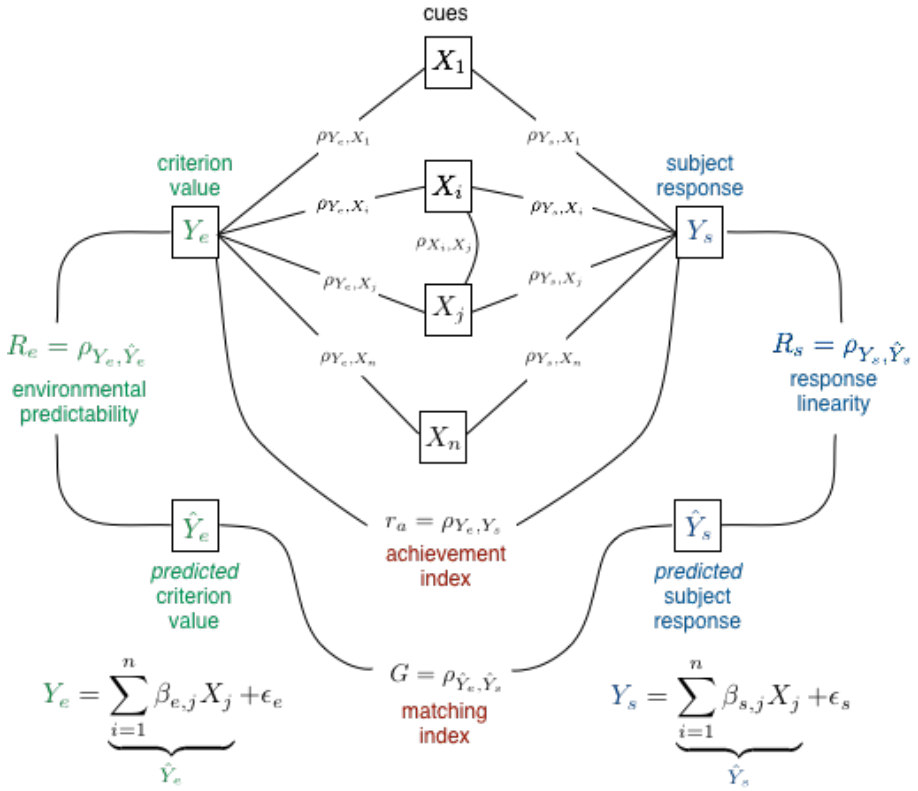


Figure 2: Brunswik's Lens Model

on what follows from the reclassification, which will depend on the model and the goal of inquiry (Section 8). If we were using the lens model to understand the ecological validity of an organism's judgment, then reclassifying ε_s as an environmental constraint would only introduce confusion; If instead our focus was to distinguish between behavior that is subject to choice and behavior that is precluded from choice, then the proposed reclassification may herald clarity—but then we would surely abandon the lens model for something else, or in any case would no longer be referring to the parameter ε_s in Figure 2.

Finally, it should be noted that the lens model, like nearly all linear models used to represent human judgment and decision-making, does not scale well as a descriptive model. In multi-cue decision-making tasks involving more than three cues, people often turn to simplifying heuristics due to the complications involved in performing the necessary calculations (Section 2.1; see also Section 4). More generally, as we remarked in Section 2.3, linear models involve calculating trade-offs that are difficult for people to perform. Lastly, the supposition that the environment is linear is a strong modeling assumption. Quite apart from the difficulties that arise for humans to execute the necessary computations, it becomes theoretically more difficult to justify model selection decisions as the number of features increases. The matching index G is a goodness-of-fit measure, but goodness-of-fit tests and residual analysis begin to lead to misleading conclusions for models with as five or more dimensions. Modern machine learning techniques for supervised learning get around this limitation by focusing on analogues of the achievement index, construct predictive hypotheses purely instrumentally, and dispense with matching altogether (Wheeler 2017).

3.3 Rational Analysis

Rational analysis is a methodology applied in cognitive science and biology to explain why a cognitive system or organism engages in a particular behavior by appealing to the presumed goals of the organism, the adaptive pressures of its environment, and the organism's computational limitations. Once an organism's goals are identified, the adaptive pressures of its environment specified, and the computa-

tional limitations are accounted for, an optimal solution under those conditions is derived to explain why a behavior that is otherwise ineffective may nevertheless be effective in achieving that goal under those conditions (Marr 1982; Anderson 1991; Oaksford and Chater 1994; Palmer 1999). Rational analyses are typically formulated independently of the cognitive processes or biological mechanisms that explain how an organism realizes a behavior.

One theme to emerge from the rational analysis literature that has influenced bounded rationality is the study of memory (Anderson and Schooler 1991). For instance, given the statistical features of our environment, and the sorts of goals we typically pursue, forgetting is an advantage rather than a liability (Schooler and Hertwig 2005). Memory traces vary in their likelihood of being used, so the memory system will try to make readily available those memories which are most likely to be useful. This is a rational analysis style argument, which is a common feature of the Bayesian turn in cognitive psychology (Oaksford and Chater 2007; Friston 2010). More generally, spacial arrangements of objectives in the environment can simplify perception, choice, and the internal computation necessary for producing an effective solution (Kirsch 1995). Compare this view to the discussion of recency or availability effects distorting subjective probability estimates in Section 2.4.

Rational analyses separate the goal of behavior from the mechanisms that cause behavior. Thus, when an organism's observed behavior in an environment does not agree with the behavior prescribed by a rational analysis for that environment, there are traditionally three responses. One strategy is to change the specifications of the problem, by introducing an intermediate step or changing the goal altogether, or altering the environmental constraints, et cetera (Anderson and Schooler 1991; Oaksford and Chater 1994). Another strategy is to argue that mechanisms matter after all, so details of human psychology are taken into an alternative account (Newell and Simon 1972; Gigerenzer, Todd, and Group 1999; Todd, Gigerenzer, and Group 2012). A third option is to enrich rational analysis by incorporating computational mechanisms directly into the model (Russell and Subramanian 1995; Chater 2014). Lewis, Howes, and Singh, for instance, propose to construct theories of rationality from (i) structural features of the task environment; (ii) the bounded machine the decision-process will run on, about which they consider four different classes of computational resources that may be available to an agent; and (iii) a utility function to specify the goal, numerically, so as to supply an objective function against which to score outcomes (Lewis, Howes, and Singh 2014).

3.4 Cultural Adaptation

So far we have considered theories and models which emphasize an individual organism and its surrounding environment, which is typically understood to be either the physical environment or, if social, modeled as if it were the physical environment. And we considered whether some features commonly understood to be behavioral constraints ought to be instead classified as environmental affordances.

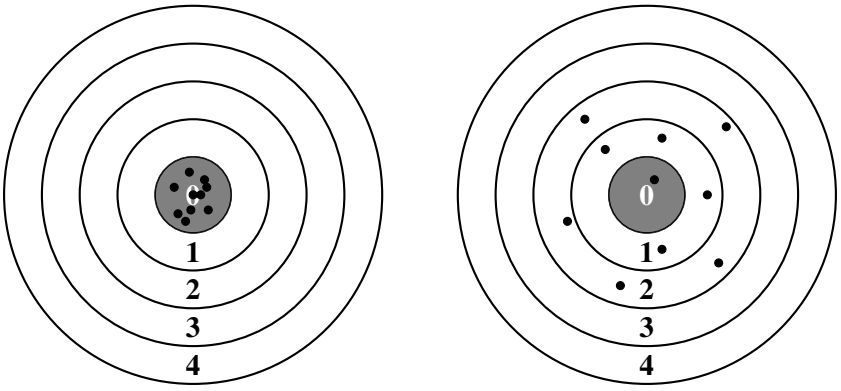
Yet people and their responses to the world are also part of each person's environment. Boyd and Richardson argue that human societies ought to be viewed as an adaptive environment, which in turn has consequences for how individual behavior is evaluated. Human societies contain a large reservoir of information that is preserved through generations and expanded upon, despite limited, imperfect learning by the members of human societies. *Imitation,* which is a common strategy in humans, including preverbal infants (Gergely, Bekkering, and Király 2002), is central to cultural transmission (Boyd and Richerson 2005) and the emergence of social norms (Bicchieri and Muldoon 2014). In our environment, only a few individuals with an interest in improving on the folk lore are necessary to nudge the culture to be adaptive. The main advantage that human societies have over other groups of social animals, this argument runs, is that cultural adaptation is much faster than genetic adaptation (Bowles and Gintis 2011). On this view, human psychology evolved to facilitate speedy adaptation. Natural selection did not equip our large-brained ancestors with rigid behavior, but instead selected for brains that allowed then to modify their behavior adaptively in response to their environment (Barkow, Cosmides, and Tooby 1992).

But if human psychology evolved to facility fast social learning, it comes at the cost of human credulity. To have speedy adaptation through imitation of social norms and human behavior, the risk is

the adoption of maladaptive norms or stupid behavior.

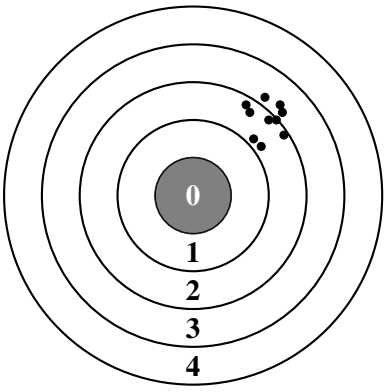
4 The Bias-Variance Trade-off

The *bias-variance trade-off* refers to a particular decomposition of overall prediction error for an estimator into its central tendency (bias) and dispersion (variance). Sometimes overall error can be reduced by increasing bias in order to reduce variance, or vice versa, effectively trading an increase in one type of error to afford a comparatively larger reduction in the other. To give an intuitive example, suppose your goal is to minimize your score with respect to the following targets.

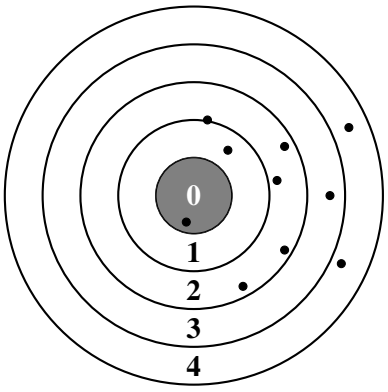


(*low bias & low variance*)

(*low bias & high variance*)



(*high bias & low variance*)



(*high bias & high variance*)

Ideally, you would prefer a procedure for delivering your "shots" that had both a low bias and low variance. Absent that, and given the choice between a low bias and high variance procedure versus a high bias and low variance procedure, you would presumably prefer the latter procedure if it returned a lower overall score than the former, which is true of the corresponding figures above. Although a decision maker's learning algorithm ideally will have low bias and low variance, in practice it is common that the reduction in one type of error yields some increase in the other. In this section we explain the conditions under which the relationship between expected squared loss of an estimator and its bias and variance holds and then remark on the role that the bias-variance trade-off plays in research on bounded rationality.

4.1 The Bias-Variance Decomposition of Mean Squared Error

Predicting the exact volume of gelato to be consumed in Rome next summer is more difficult than predicting that more gelato will be consumed next summer than next winter. For although it is a foregone conclusion that higher temperatures beget higher demand for gelato, the precise relationship between daily temperatures in Rome and *consumo di gelato* is far from certain. Modeling quantitative, predictive

relationships between random variables, such as the relationship between the temperature in Rome, X, and volume of Roman gelato consumption, Y, is the subject of *regression analysis*.

Suppose we predict that the value of Y is h. How should we evaluate whether this prediction is any good? Intuitively, the best we can do is to pick an h that is as close to Y as we can make it, one that would minimize the difference $Y - h$. If we are indifferent to the direction of our errors, viewing positive errors of a particular magnitude to be no worse than negative errors of the same magnitude, and vice versa, then a common practice is to measure the performance of h by its squared difference from Y, $(Y-h)^2$. (We are not always indifferent; consider the plight of William Tell aiming at that apple.) Finally, since the values of Y vary, we might be interested in the average value of $(Y-h)^2$ by computing its expectation, $\mathbb{E}\left[(Y-h)^2\right]$. This quantity is the *mean squared error* of h,

$$\mathrm{MSE}(h) := \mathbb{E}\left[(Y-h)^2\right].$$

Now imagine our prediction of Y is based on some data $\mathcal{D}$ about the relationship between X and Y, such as last year's daily temperatures and daily total sales of gelato in Rome. The role that this particular dataset $\mathcal{D}$ plays as opposed to some other possible data set is a detail that will figure later. For now, view our prediction of Y as some function of X, written $h(X)$. Here again we wish to pick an $h(\cdot)$ to minimize $\mathbb{E}\left[(Y-h(X))^2\right]$, but how close $h(\cdot)$ is to Y will depend on the possible values of X, which we can represent by the conditional expectation

$$\mathbb{E}\left[(Y-h(X))^2\right] := \mathbb{E}\left[\mathbb{E}\left[Y-h(X) \mid X\right]\right].$$

How then should we evaluate this conditional prediction? The same as before, only now accounting for X. For each possible value x of X, the best prediction of Y is the conditional mean, $\mathbb{E}[Y \mid X = x]$. The *regression function* of Y on X, $r(x)$, gives the optimal value of Y for each value $x \in X$:

$$r(x) := \mathbb{E}[Y \mid X = x].$$

Although the regression function represents the true population value of Y given X, this function is usually unknown, typically complicated, therefore often approximated by a simplified model or learning algorithm, $h(\cdot)$.

We might restrict candidates for $h(X)$ to linear (or affine) functions of X, for instance. Yet making predictions about the value of Y with a simplified linear model, or some other simplified model, can introduce a systematic prediction error called *bias*. Bias results from a difference between the central tendency of data generated by the true model, $r(X)$ (for all $x \in X$), and the central tendency of our estimator, $\mathbb{E}[h(X)]$, written

$$\mathrm{Bias}(h(X)) := r(X) - \mathbb{E}[h(X)],$$

where any non-zero difference between the pair is interpreted as a systematically positive or systematically negative error of the estimator, $h(X)$.

Variance measures the average deviation of a random variable from its expected value. In the current setting we are comparing the predicted value $h(X)$ of Y, with respect to some data $\mathcal{D}$ about the relationship between X and Y, and the average value of $h(X)$, $\mathbb{E}[h(X)]$, which we will write

$$\mathrm{Var}(h(X)) = \mathbb{E}\left[(\mathbb{E}[h(X)] - h(X))^2\right].$$

The bias-variance decomposition of mean squared error is rooted in frequentist statistics, where the objective is to compute an estimate $h(X)$ of the true parameter $r(X)$ with respect to data $\mathcal{D}$ about the relationship between X and Y. Here the parameter $r(X)$ characterizing the truth about Y is assumed to be fixed and the data $\mathcal{D}$ is treated as a random quantity, which is exactly the reverse of Bayesian statistics. What this means is that the data set $\mathcal{D}$ is interpreted to be one among many possible data sets of the same dimension generated by the true model, the deterministic process $r(X)$.

Following (Bishop 2006), we may derive the bias-variance decomposition of mean squared error of h as follows. Let h refer to our estimate $h(X)$ of Y, r refer to the true value of Y, and $\mathbb{E}[h]$ the expected value of the estimate h. Then,

$$\begin{aligned}
\text{MSE}(h) &= \mathbb{E}\left[(r-h)^2\right] \\
&= \mathbb{E}\left[((r-\mathbb{E}[h])+(\mathbb{E}[h]-h))^2\right] \\
&= \mathbb{E}\left[(r-\mathbb{E}[h])^2\right] + \mathbb{E}\left((\mathbb{E}[h]-h)^2\right) + 2\mathbb{E}\left[(\mathbb{E}[h]-h)\cdot(r-\mathbb{E}[h])\right] \\
&= (r-\mathbb{E}[h])^2 + \mathbb{E}\left[(\mathbb{E}[h]-h)^2\right] + 0 \\
&= \text{B}(h)^2 \; + \; \text{Var}(h)
\end{aligned}$$

where the term $2\mathbb{E}\left[(\mathbb{E}[h]-h)\cdot(r-\mathbb{E}[h])\right]$ is zero, since

$$\begin{aligned}
\mathbb{E}\left[(\mathbb{E}[h]-h)\cdot(r-\mathbb{E}[h])\right] &= \left(\mathbb{E}[r\cdot\mathbb{E}[h]] - \mathbb{E}\left[\mathbb{E}[h]^2\right] - \mathbb{E}[h\cdot r] + \mathbb{E}[h\cdot\mathbb{E}[h]]\right) \\
&= r\cdot\mathbb{E}[h] - \mathbb{E}[h]^2 - r\cdot\mathbb{E}[h] + \mathbb{E}[h]^2 \qquad (2) \\
&= 0.
\end{aligned}$$

Note that the frequentist assumption that r is a deterministic process is necessary for the derivation to go through; for if r were a random quantity, the reduction of $\mathbb{E}[r\cdot\mathbb{E}[h]]$ to $r\cdot\mathbb{E}[h]$ in line (2) would be invalid.

One last detail that we have skipped over is the prediction error of $h(X)$ due to noise, N, which occurs independent of the model/learning algorithm used. Thus, the full bias-variance decomposition of the mean-squared error of an estimate h is the sum of the bias (squared), variance, and irreducible error:

$$\text{MSE}(h) \; = \; \text{B}(h)^2 \; + \; \text{Var}(h) \; + \; N \qquad (3)$$

4.2 Bounded Rationality and Bias-Variance Generalized

Intuitively, the bias-variance decomposition brings to light a trade-off between two extreme approaches to making a prediction. At one extreme, you might adopt as an estimator a constant function which produces the same answer no matter what data you see. Suppose 7 is your lucky number and your estimator's prediction, $h(X) = 7$. Then the variance of $h(\cdot)$ would be zero, since its prediction is always the same. The bias of your estimator, however, will be very large. In other words, your lucky number 7 model will massively *underfit* your data.

At the other extreme, suppose you aim to make your bias error zero. This occurs just when the predicted value of Y and the actual value of Y are identical, that is, $h(x_i) = y_i$, for every (x_i, y_i). Since you are presumed to not know the true function $r(X)$ but instead only see a sample of data from the true model, $\mathcal{D}$, it is from this sample that you will aspire to construct an estimator that generalizes to accurately predict examples outside your training data $\mathcal{D}$. Yet if you were to fit $h_{\mathcal{D}}(X)$ perfectly to $\mathcal{D}$, then the variance of your estimator will be very high, since a different data set $\mathcal{D}'$ from the true model is not, by definition, identical to $\mathcal{D}$. How different is $\mathcal{D}'$ to $\mathcal{D}$? The variation from one data set to another among all the possible data sets is the variance or irreducible noise of the data generated by the true model, which may be considerable. Therefore, in this zero-bias case your model will massively *overfit* your data.

The bias-variance trade-off therefore concerns the question of how complex a model ought to be to make reasonably accurate predictions on unseen or out-of-sample examples. The problem is to strike a balance between an underfitting model, which erroneously ignores available information about the true function r, and an overfitting model, which erroneously includes information that is noise and thereby gives misleading information about the true function r.

One thing that human cognitive systems do very well is to generalize from a limited number of examples. The difference between humans and machines is particularly striking when we compare how

humans learn a complicated skill, such as driving a car, from how a machine learning system learns the same task. As harrowing an experience it is to teach a teenager how to drive a car, they do not need to crash into a utility pole 10,000 times to learn that utility poles are not traversable. What teenagers learn as children about the world through play and observing other people drive lends to them an understanding that that utility poles are to be steered around, a piece of commonsense that our current machine learning systems do not have but must learn from scratch on a case-by-case basis. We, unlike our machines, have a remarkable capacity to transfer what we learn from one domain to another domain, a capacity fueled in part by our curiosity (Kidd and Hayden 2015).

Viewed from the perspective of the bias-variance trade-off, the ability to make accurate predictions from sparse data suggests that variance is the dominant source of error but that our cognitive system often manages to keep these errors within reasonable limits (Gigerenzer and Brighton 2009). Indeed, Gigerenzer and Brighton make a stronger argument, stating that " the bias-variance dilemma shows formally why a mind can be better off with an adaptive toolbox of biased, specialized heuristics" (Gigerenzer and Brighton 2009, p. 120); see also (Section 7.2). However, the bias-variance decomposition is a decomposition of squared loss, which means that the decomposition above depends on how total error (loss) is measured. There are many loss functions, however, depending on the type of inference one is making along with the stakes in making it. If one were to use a 0-1 loss function, for example, where all non-zero errors are treated equally—meaning that "a miss as good as a mile"—the decomposition above breaks down. In fact, for 0-1 loss, bias and variance combine multiplicatively (Friedman 1997)! A generalization of the bias-variance decomposition that applies to a variety of loss functions $\mathrm{L}(\cdot)$, including 0-1 loss, has been offered by (Domingos 2000),

$$\mathrm{L}(h) \;=\; \mathrm{B}(h)^2 \;+\; \beta_1 \mathrm{Var}(h) \;+\; \beta_2 \mathrm{N}$$

where the original bias-variance decomposition, Equation 3, appears as a special case, namely when $\mathrm{L}(h) = \mathrm{MSE}(h)$ and $\beta_1 = \beta_2 = 1$.

5 Better with Bounds

Our discussion of improper linear models (Section 2.3) mentioned a model that often comes surprisingly close to approximating a proper linear model, and our discussion of the bias-variance decomposition (Section 4.2) referred to conjectures about how cognitive systems might manage to make accurate predictions with very little data . In this section we review examples of models which deviate from the normative standards of global rationality yet yield markedly *improved* outcomes—sometimes even yielding results which are impossible under the conditions of global rationality. Thus, in this section we will survey examples from the *statistics of small samples* and *game theory* which point to demonstrable advantages to deviating from global rationality.

5.1 Homo Statisticus and Small Samples

In a review of experimental results assessing human statistical reasoning published in the late 1960s that took stock of research conducted after psychology's full embrace of statistical research methods (Section 2.3), Petersen and Beach argued that the normative standard of probability theory and statistical optimization methods were "a good first approximation for a psychological theory of inference" (Peterson and Beach 1967, p. 42). Petersen and Beach's view that humans were *intuitive statisticians* that closely approximate the ideal standards of *homo statisticus* fit into a broader consensus at that time about the close fit between the normative standards of logic and intelligent behavior (Newell and Simon 1956; Newell and Simon 1976). The assumption that human judgment and decision-making closely approximates normative theories of probability and logic would later be challenged by experimental results by Kahneman and Tversky, and the biases and heuristics program more generally (Section 7.1).

Among Kahneman and Tversky's earliest findings was that people tend to make statistical inferences from samples that are too small, even when given the opportunity to control the sampling procedure.

Kahneman and Tversky attributed this effect to a systematic failure of people to appreciate the biases that attend small samples, although Hertwig and others have offered evidence that samples drawn from a single population are close to the known limits to working memory (Hertwig, Barron, Weber, and Erev 2004).

Overconfidence can be understood as an artifact of small samples. The *Naïve Sampling Model* (Juslin, Winman, and Hannson 2007) assumes that agents base judgments on a small sample retrieved from long-term memory at the moment a judgment is called for, even when there are a variety of other methods available to the agent. This model presumes that people are naïve statisticians (Fiedler and Juslin 2006) who assume, sometimes falsely, that samples are representative of the target population of interest and that sample properties can be used directly to yield accurate estimates of a population. The idea is that when sample properties are uncritically taken as estimators of population parameters a reasonably accurate probability judgment can be made with overconfidence, even if the samples are unbiased, accurately represented, and correctly processed by the cognitive mechanisms of the agent. When sample sizes are restricted, these effects are amplified.

However, sometimes effective behavior is aided by inaccurate judgments or cognitively adaptive illusions (Howe 2011). The statistical properties of small samples are a case in point. One feature of small samples is that correlations are amplified, making them easier to detect (Kareev 1995). This fact about small samples, when combined with the known limits to human short-term memory, suggests that our working-memory limits may be an adaptive response to our environment that we exploit at different stages in our lives. Adult short-term working memory is limited to seven items, plus or minus two. For correlations of 0.5 and higher, Kareev demonstrates that sample sizes between five and nine are most likely to yield a sample correlation that is greater than the true correlation in the population (Kareev 2000), making those correlations nevertheless easier to detect. Furthermore, children's short-term memories are even more restricted than adults, thus making correlations in the environment that much easier to detect. Of course, there is no free lunch: this small-sample effect comes at the cost of inflating estimates of the true correlation coefficients and admitting a higher rate of false positives (Juslin and Olsson 2005). However, in many contexts, including child development, the cost of error arising from under-sampling may be more than compensated by the benefits from simplifying choice (Hertwig and Pleskac 2008) and accelerating learning. In the spirit of Brunswik's argument for representative experimental design (Section 3.2), a growing body of literature cautions that the bulk of experiments on adaptive decision-making are performed in highly simplified environments that differ in important respects from the natural world in which human beings make decisions (Fawcett, Fallenstein, Higginson, Houston, Mallpress, Trimmer, and McNamara 2014). In response, Houston, MacNamara and colleagues argue, we should incorporate more environmental complexity in our models.

5.2 Game Theory

Pro-social behavior, such as cooperation, is challenging to explain. Evolutionary game theory predicts that individuals will forgo a public good and that individual utility maximization will win over collective cooperation. Even though this outcome is often seen in economic experiments, in broader society cooperative behavior is pervasive (Bowles and Gintis 2011). Why? The traditional evolutionary explanations of human cooperation in terms of *reputation*, *reciprocation*, and *retribution* (Trivers 1971; Alexander 1987), are unsatisfactory because they do not uniquely explain why cooperation is a stable behavior. If a group punishes individuals for failing to perform a behavior, and the punishment costs exceed the benefit of doing that behavior, then this behavior will become stable regardless of its social benefits. Anti-social norms arguably take root by precisely the same mechanisms (Bicchieri and Muldoon 2014). Although reputation, reciprocation, and retribution may explain how large-scale cooperation is sustained in human societies, it does not explain how the behavior emerged (Boyd and Richerson 2005). Furthermore, cooperation is observed in microorganism (Damore and Gore 2012), which suggests that much simpler mechanisms are sufficient for the emergence of cooperative behavior.

Whereas the 1970s saw a broader realization of the advantages of improper models to yield results that were often good enough (Section 2.3), the 1980s and 1990s witnessed a series of results involving

improper models yielding results that were strictly better than what was prescribed by the corresponding proper model. In the early 1980s Robert Axelrod held a tournament to empirically test which among a collection of strategies for playing iterations of the prisoner's dilemma performed best in a round-robin competition. The winner was a simple reciprocal altruism strategy called *tit-for-tat* (Rapoport and Chammah 1965), which simply starts off each game cooperating then, on each successive round, copies the strategy the opposing player played in the previous round. So, if your opponent cooperated in this round, then you will cooperate on the next round; and if your opponent defected this round, then you will defect the next. Subsequent tournaments have shown that tit-for-tat is remarkably robust against much more sophisticated alternatives (Axelrod 1984). For example, even a rational utility maximizing player playing against an opponent who only plays tit-for-tat (i.e., will play tit-for-tat no matter whom he faces) must adapt and play tit-for-tat—or a strategy very close to it (Kreps, Milgrom, Roberts, and Wilson 1982).

Since tit-for-tat is a very simple strategy, computationally, one can begin to explore a notion of rationality that emerges in a group of boundedly rational agents and even see evidence of those bounds contributing to the emergence of pro-social norms. Rubinstein (Rubinstein 1986) studied finite automata which play repeated prisoner's dilemmas and whose aims are to maximize average payoff while minimizing the number of states of a machine. Finite automata capture regular languages, the lowest-level of the Chomsky-hierarchy, thus model a type of boundedly rational agents. Solutions are a pair of machines in which the choice of the machine is optimal for each player at every stage of the game. In an evolutionary interpretation of repeated games, each iteration of Rubinstein's can be seen as successive generations of agents. This approach is in contrast to Neyman's study of players of repeated games who can only play mixtures of pure strategies that can be programmed on finite automata, where the number of states that are available is an exogenous variable whose value is fixed by the modeler. In Neyman's model, each generation plays the entire game and thus traits connected to reputation can arise (Neyman 1985). More generally, although cooperation is impossible for infinitely repeated prisoner's dilemmas, for finitely repeated prisoner's dilemmas, a cooperative equilibrium exists for finite automata players whose number of states is less than exponential in the number of rounds of the game (Papadimitriou and Yannakakis 1994; Ho 1996). The demands on memory may exceed the psychological capacities of people, however, even for simple strategies like tit-for-tat played by a moderately sized group of players (Stevens, Volstorf, Schooler, and Rieskamp 2011). These theoretical models showing a number of simple paths to pro-social behavior may not, on their own, be simple enough to offer plausible process models for cooperation.

On the heels of work on the effects of time (finite iteration versus infinite iteration) and memory/cognitive ability (finite state automata versus Turing machines), attention soon turned to environmental constraints. Nowak and May looked at the spatial distribution on a two-dimensional grid of 'cooperators' and 'defectors' in iterated prisoner's dilemmas and found cooperation to emerge among players without memories or strategic foresight (Nowak and May 1992). This work led to the study of *network topology* as a factor in social behavior (Jackson 2010), including social norms (Bicchieri 2005; Alexander 2007), signaling (Skyrms 2003), and wisdom of crowd effects (Golub and Jackson 2010). When social ties in a network follow a scale-free distribution, the resulting diversity in the number and size of public-goods games is found to promote cooperation, which contributes to explaining the emergence of cooperation in communities without mechanisms for reputation and punishment (Santos, Santos, and Pacheco 2008).

But, perhaps the simplest case for bounded rationality are examples of agents achieving a desirable goal without any deliberation at all. Insects, flowers, and even bacteria exhibit evolutionary stable strategies (Maynard Smith 1982), effectively arriving at Nash equilibria in strategic normal form games. If we imagine two species interacting with one another, say honey bees (*Apis mellifera*) and a species of flower, each interaction between a bee and a flower has some bearing on the fitness of each species, where fitness is defined as the expected number of offspring. There is an incremental payoff to bees and flowers, possibly negative, after each interaction, and the payoffs are determined by the genetic endowments of bees and flowers each. The point is that there is no choice exhibited by these organisms nor

in the models; the process itself selects the traits. The agents have no foresight. There are no strategies that the players themselves choose. The process is entirely mechanical. What emerges in this setting are *evolutionary dynamics*, a form of bounded rationality without foresight.

Of course, any improper model can misfire. A rule of thumb shared by people the world-over is to not let other people take advantage of them. While this rule works most of the time, it misfires in the *ultimatum game* (Güth, Schmittberger, and Schwarze 1982). The ultimatum game is a two-player game in which one player, endowed with a sum of money, is given the task of splitting the sum with another player who may either accept the offer—in which case the pot is accordingly split between the two players—or rejected, in which case both players receiving nothing. People receiving offers of 30 percent or less of the pot are often observed to reject the offer, even when players are anonymous and therefore would not suffer the consequences of a negative reputation signal associated with accepting a very low offer. In such cases, one might reasonably argue that no proposed split is worse than the status quo of zero, so people ought to accept whatever they are offered.

5.3 Less is More Effects

Simon's remark that people satisfice when they haven't the wits to maximize (Simon 1957, p. xxiv) points to a common assumption, that there is a trade-off between effort and accuracy (Section 2.1). Because the rules of global rationality are expensive to operate (Good 1952, §7(i)), people will trade a loss in accuracy for gains in cognitive efficiency (Payne, Bettman, and Johnson 1988). The methodology of rational analysis (Section 3.3) likewise appeals to this trade-off.

The results surveyed in Section 5.2 caution against blindly endorsing the accuracy-effort trade-off as universal, a point that has been pressed in the defense of heuristics as reasonable models for decision-making (Katsikopoulos 2010; Hogarth 2012).

Simple heuristics like *Tallying*, which is a type of improper linear model (Section 2.3), and *Take-the-best* (Section 7.2), when tested against linear regression on many data sets, have been both found to outperform linear regression on out-of-sample prediction tasks, particularly when the training-sample size is low (Czerlinski, Gigerenzer, and Goldstein 1999; Rieskamp and Dieckmann 2012).

6 Aumann's Five Arguments and One More

Aumann advanced five arguments for bounded rationality, which we paraphrase here (Aumann 1997).

1. Even in very simple decision problems, most economic agents are not (deliberate) maximizers. People do not scan the choice set and consciously pick a maximal element from it.

2. Even if economic agents aspired to pick a maximal element from a choice set, performing such maximizations are typically difficult and most people are unable to do so in practice.

3. Experiments indicate that people fail to satisfy the basic assumptions of rational decision theory.

4. Experiments indicate that the conclusions of rational analysis (broadly construed to include rational decision theory) do not match observed behavior.

5. Some conclusions of rational analysis appear normatively unreasonable.

In the previous sections we covered the origins of each of Aumann's arguments. Here we briefly review each, highlighting material in other sections under this context.

The first argument, that people are not deliberate maximizers, was a working hypothesis of Simon's, who maintained that people tend to satisfice rather than maximize (Section 2.2). Kahneman and Tversky gathered evidence for the reflection effect in estimating the value of options, which is the reason for reference points in prospect theory (Section 2.4) and analogous properties within rank-dependent utility theory more generally (Sections 1.2 and 2.4). Gigerenzer's and Hertwig's groups at the Max Planck

Institute for Human Development both study the algorithmic structure of simple heuristics and the adaptive psychological mechanisms which explain their adoption and effectiveness; both of their research programs start from the assumption that expected utility theory is not the right basis for a descriptive theory of judgment and decision-making (Sections 3, 5.3, and 7.2).

The second argument, that people are often unable to maximize even if they aspire to, was made by Simon and Good, among others, and later by Kahneman and Tversky. Simon's remarks about the complexity of Γ-maxmin reasoning in working out the end-game moves in chess (Section 2.2) is one of many examples he used over the span of his career, starting before his seminal papers on bounded rationality in the 1950s. The *biases and heuristics* program spurred by Tversky and Kahneman's work in the late 1960s and 1970s (Section 7.1) launched the systematic study of when and why people's judgments deviate from the normative standards of expected utility theory and logical consistency.

The third argument, that experiments indicate that people fail to satisfy the basic assumptions of expected utility theory, was known from early on and emphasized by the very authors who formulated and refined the homo economicus hypothesis (Section 1) and whose names are associated with the mathematical foundations. We highlighted an extended quote from Savage in Section 1.3, but could mention as well a discussion of the theory's limitations by de Finetti and Savage (de Finetti and Savage 1962), and even a closer reading of the canonical monographs of each, namely (Savage 1954) and (de Finetti 1974). A further consideration, which we discussed in Section 1.3 is the demand of *logical omniscience* in expected utility theory and nearly all axiomatic variants.

The fourth argument, regarding the differences between the predictions of rational analysis and observed behavior, we addressed in discussions of Brunswik's notion of ecological validity (Section 3.2) and the traditional responses to these observations by rational analysis (Section 3.3). The fifth argument, that some of the conclusions of rational analysis do not agree with a reasonable normative standard, was touched on in Sections 1.2, 1.3, and the subject of Section 5.

Implicit in Aumann's first four arguments is the notion that global rationality (Section 2) is a reasonable normative standard but problematic for descriptive theories of human judgment and decision-making (Section 8). Even the literature standing behind Aumann's 5th argument, namely that there are problems with expected utility theory as a normative standard, nevertheless typically address those shortcomings through modifications to, or extensions of, the underlying mathematical theory (Section 1.2). This broad commitment to optimization methods, dominance reasoning, and logical consistency as bedrock normative principles is behind approaches that view bounded rationality as *optimization under constraints*.

> Boundedly rational procedures are in fact fully optimal procedures when one takes account of the cost of computation in addition to the benefits and costs inherent in the problem as originally posed (Arrow 2004).

For a majority of researchers across disciplines, bounded rationality is identified with some form of optimization problem under constraints.

Gerd Gigerenzer is among the most prominent and vocal critics of the role that optimization methods and logical consistency plays in commonplace normative standards for human rationality (Gigerenzer and Brighton 2009), especially the role those standards play in Kahneman and Tversky's biases and heuristics program (Kahneman and Tversky 1996; Gigerenzer 1996). We turn to this debate next, in Section 7.

7 Two Schools of Heuristics

Heuristics are simple rules of thumb for rendering a judgment or making a decision. Some examples that we have seen thus far include Simon's satisficing, Dawes's improper linear models, Rapoport's tit-for-tat, imitation, and several effects observed by Kahneman and Tversky in our discussion of prospect theory.

There are nevertheless two views on heuristics that are roughly identified with the research traditions associated with Kahneman and Tversky's *biases and heuristics* program and Gigerenzer's *fast and*

frugal heuristics program, respectively. A central dispute between these two research programs is the appropriate normative standard for judging human behavior (Vranas 2000). According to Gigerenzer, the biases and heuristics program mistakenly classifies all biases as errors (Gigerenzer, Todd, and Group 1999; Gigerenzer and Brighton 2009) despite evidence pointing to some biases in human psychology being adaptive. In contrast, in a rare exchange with a critic, Kahneman and Tversky maintain that the dispute is merely terminological (Kahneman and Tversky 1996; Gigerenzer 1996).

In this section, briefly survey each of these two schools. Our aim is to give a characterization of each research program rather than an exhaustive overview.

7.1 Biases and Heuristics

Beginning in the 1970s, Kahneman and Tversky conducted a series of experiments showing various ways that human participants' responses to decision tasks deviate from answers purportedly derived from the appropriate normative standards (Sections 2.4 and 5.1). These deviations were given names, such as *availability* (Tversky and Kahneman 1973), *representativeness*, and *anchoring* (Tversky and Kahneman 1974). The set of cognitive biases now numbers into the hundreds, although some are minor variants of other well-known effects, such as "The IKEA effect" (Norton, Mochon, and Ariely 2012) being a version of the well-known endowment effect (Section 1.2). Nevertheless, core effects studied by the biases and heuristics program, particularly those underpinning prospect theory (Section 2.4), are entrenched in cognitive psychology (Kahneman, Slovic, and Tversky 1982).

An example of a probability judgment task is Kahneman and Tversky's Taxi-cab problem, which purports to show that subjects neglect base rates.

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (i) 85% of the cabs in the city are Green and 15% are Blue.
>
> (ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.
>
> Question: What is the probability that the cab involved in the accident was Blue rather than Green? (Tversky and Kahneman 1977, §3-3).

Continuing, Kahneman and Tverskey report that several hundred subjects have been given slight variations of this question and for all versions the modal and median responses was 0.8, instead of the correct answer of $12/29$ (≈ 0.41). "Thus, the intuitive judgment of probability coincides with the credibility of the witness and ignores the relevant base-rate, i.e., the relative frequency of Green and Blue cabs" (Tversky and Kahneman 1977, §3-3).

Critical responses to results of this kind fall into three broad categories. The first types of reply is to argue that the experimenters, rather than the subjects, are in error (Cohen 1981). In the Taxi-cab problem, arguably Bayes sides with the folk (Levi 1983) or, alternatively, is inconclusive because the normative standard of the experimenter and the presumed normative standard of the subject requires a theory of witness testimony, neither of which is specified (Birnbaum 1979). Other cognitive biases have been ensnared in the replication crises, such as *implicit bias* (Oswald, Mitchell, Blanton, Jaccard, and Tellock 2013; Forscher, Lai, Axt, Ebersole, Herman, Devine, and Nosek 2017) and *social priming* (Doyen, Klein, Pichton, and Cleeremans 2012; Kahneman 2017).

The second response is to argue that there is an important difference between identifying a normative standard for combining probabilistic information and applying it across a range of cases (Section 8.2), and it is difficult in practice to determine that a decision-maker is representing the task in the manner that the experimenters intend (Koehler 1996). Observed behavior that appears to be boundedly rational

or even irrational may result from a difference between the intended specification of a problem and the actual problem subjects face.

For example, consider the systematic biases in people's perception of randomness reported in some of Kahneman and Tversky's earliest work (Kahneman and Tversky 1972). For sequences of flips of a fair coin, people expect to see, even for small samples, a roughly-equal number heads and tails and alternation rates between heads and tails that are slightly higher than long-run averages (Bar Hillel and Wagenaar 1991). This effect is thought to explain the *gambler's fallacy*, the false belief that a run of heads from an i.i.d. sequence of fair coin tosses will make the next flip more likely to land tails. Hahn and Warren argue that the limited nature of people's experiences with random sequences is a better explanation than to view them as cognitive deficiencies. Specifically, people only ever experience finite sequence of outputs from a randomizer, such as a sequence of fair coin tosses, and the limits to their memory (Section 5.1) of past outcomes in a sequence will mean that not all possible sequences of a given length with appear to them with equal probability. Therefore, there is a psychologically plausible interpretation of the question, "*is it more likely to see* HHHT *than* HHHH *from flips of a fair coin?*", for which the correct answer is, "Yes" (Hahn and Warren 2009). If the gambler's fallacy boils down to a failure to distinguish between sampling with and without replacement, Hahn and Warren's point is that our intuitive statistical abilities acquired through experience along is unable to make the distinction between these two sampling methods. Analytical reasoning is necessary.

Consider also the risky-choice framing effect that was mentioned briefly in Section 2.4. An example is the Asian disease example,

(a) If program *A* is adopted, 200 people will be saved.

(b) If program *B* is adopted, there is a $1/3$ probability that 600 people will be saved, and a $2/3$ probability that no people will be saved (Tversky and Kahneman 1981, p. 453).

Tversky and Kahneman report that a majority of respondents (72 percent) chose option (a), whereas a majority of respondents (78 percent) shown an equivalent reformulation of the problem in terms of the number of people who would die rather than survive chose (b). A meta-analysis of subsequent experiments has shown that the framing condition accounts for most of the variance, but it also reveals no linear combination of formally specified predictors that are used in prospect theory, cumulative prospect theory, and Markowitz's utility theory, suffices to capture this framing effect (Kühberger, Schulte-Mecklenbeck, and Perner 1999). Furthermore, the use of an indicative conditional in this and other experiments to express the consequences is also not (currently) adequately understood. Experimental evidence collected about how people's judgments change when learning an indicative conditional, while straight-forward and intuitive, cannot be accommodated by existing theoretical frameworks for conditionals (Collins, Krzyż, Hartmann, Wheeler, and Hahn 2018).

The point to this second line of criticism is not that people's responses are at variance with the correct normative standard but rather that the explanation for why they are at variance will matter not only for assessing the rationality of people but what prescriptive interventions ought to be taken to counter the error. It is rash to conclude that people, rather than the peculiarities of the task or the theoretical tools available to us at the moment, are in error.

Lastly, the third type of response is to accept the experimental results but challenge the claim that they are generalizable. In a controlled replication of Kahneman and Tversky's lawyer-engineer example (Tversky and Kahneman 1977), for example, a crucial assumption is whether the descriptions of the individuals were drawn at random, which was tested by having subjects draw blindly from an urn (Gigerenzer, Hell, and Blank 1988). Under these conditions, base-rate neglect disappeared. In response to the Linda example (Tversky and Kahneman 1983), rephrasing the example in terms of *which alternative is more frequent* rather than *which alternative is more probable* reduces occurrences of the conjunction fallacy among subjects from 77% to 27% (Fiedler 1988). More generally, a majority of people presented with the Linda example appear to interpret 'probability' nonmathematically but switch to a mathematical interpretation when asked for frequency judgments (Hertwig and Gigerenzer 1999).

Ralph Hertwig and colleagues have since noted a variety of other effects involving probability judgments to diminish or disappear when subjects are permitted to learn the probabilities through sampling, suggesting that people are better adapted to making a *decision by experience* of the relevant probabilities as opposed to making a decision by their *description* (Hertwig, Barron, Weber, and Erev 2004).

7.2 Fast and Frugal Heuristics

The Fast and Frugal school and the Biases and Heuristics school both agree that heuristics are biased. Where they disagree, and disagree sharply, is whether those biases are necessarily a sign of irrationality. For the Fast and Frugal program the question is under what environmental conditions, if any, does a particular heuristic perform effectively. If the heuristic's structural bias is well-suited to the task environment, then the bias of that heuristic may be an advantage for making accurate judgments rather than a liability (Section 4). We saw this adaptive strategy before in our discussion of Brunswik's lens model (Section 3.2), although there the bias in the model was to assume that both the environment and the subject's responses were linear. The aim of the Fast and Frugal program is to adapt this Brunswikian strategy to a variety of improper models.

This general goal of the Fast and Frugal program leads to a second difference between the two schools. Because the Fast and Frugal program aims to specify the conditions under which a heuristic will lead to better outcomes than competing models, heuristics are treated as algorithmic models of decision-making rather than descriptions of errant effects; heuristics are themselves objects of study. To that end, all heuristics in the fast and frugal tradition are conceived to have three components: (i) a search rule, (ii) a stopping rule, and (iii) a decision rule. For example, *Take-the-Best* (Gigerenzer and Goldstein 1996), is a heuristic applied to binary, forced-choice problems. Specifically, the task is to pick the correct option according to an external criterion, such as correctly picking which of a pair of cities has a larger population, based on cue information that is available to the decision-maker, such as whether she has heard of one city but not the other, whether one city is known to have a football franchise in the professional league, et cetera. Based on data sets, one can compute the predictive validity of different cues, and thus derive their weights. Take-the-Best then has the following structure: *Search rule:* Look up the cue with the highest cue-validity; *Stopping rule:* If the pair of objects have different cue values, that is, one is positive and the other negative, stop the search. If the cue values are the same, continue searching down the cue-order; *Decision rule:* Predict that the alternative with the positive cue value has the higher target-criterion value. If all cues fail to discriminate, that is, if all cue values are the same, then predict the alternative randomly by a coin flip. The bias of Take-the-Best is that it ignores relevant cues. Another example is *tallying*, which is a type of improper linear model (Section 2.3). Tallying has the following structure for a binary, forced-choice task: *Search rule:* Look up cues in a random order; *Stopping rule:* After some exogenously determined m $(1 < m \leq N)$ of the N available cues are evaluated, stop the search; *Decision rule:* Predict that the alternative with the higher number of positive cue values has the higher target-criterion value. The bias in tallying is that it ignores cue weights. One can see then how models are compared to one another by how they process cues and their performance is evaluated with respect to a specified criterion for success, such as the number of correct answers to the city population task.

Because Fast and Frugal heuristics are computational models, this leads to a third difference between the two schools. Kahneman endorses the System I and System II theory of cognition (Stanovich and West 2000). Furthermore, Kahneman classifies heuristics as fast, intuitive, and non-deliberative System I thinking. Gigerenzer, by contrast, does not endorse the System I and System II hypothesis, thus rejects classifying heuristics as, necessarily, non-deliberative cognitive processes. Because heuristics are computational models in the Fast and Frugal program, in principle each may be used deliberatively by a decision-maker or used by a decision-modeler to explain or predict a decision-maker's non-deliberative behavior. The Linear Optical Trajectory (LOT) heuristic (McBeath, Shaffer, and Kaiser 1995) that baseball players use intuitively, without deliberation, to catch fly balls, and which some animals appear to use to intercept prey, is the same heuristic that the "Miracle on the Hudson" airline pilots used deliberatively to infer that they could not reach an airport runway and decided instead to land their crippled plane in

the Hudson River.

Here are a list of heuristics studied in the Fast and Frugal program (Gigerenzer, Hertwig, and Pachur 2011), along with an informal description for each along with historical and selected contemporary references.

Imitation. *People have a strong tendency to imitate the successful members of their communities* (Henrich and Gil-White 2001). "If some man in a tribe ... invented a new snare or weapon, or other means of attack or defense, the plainest self-interest, without the assistance of much reasoning power, would prompt other members to imitate him" (Darwin 1871, p. 155). Imitation is presumed to be fundamental to the speed of cultural adaptation including the adoption of social norms (Section 3.4).

Preferential Attachment. *When given the choice to form a new connection to someone, pick the individual with the most connections to others* (Yule 1911; Simon 1955b; Barabási and Albert 1999).

Default rules. *If there is an applicable default rule, and no apparent reason for you to do otherwise, follow the rule.* (Fisher 1936; Reiter 1980; Wheeler 2004; Thaler and Sustein 2008).

Satisficing. *Search available options and choose the first one that exceeds your aspiration level.* (Simon 1955a; Hutchinson, Fanselow, and Todd 2012).

Tallying. *To estimate a target criterion, rather than estimate the weights of available cues, instead count the number of positive instances* (Dawes 1979; Dana and Dawes 2004).

One-bounce Rule (Hey's Rule B). *Have at least two searches for an option. Stop if a price quote is larger than the previous quote.* The one-bounce rule plays "winning-streaks" by continuing search while you keep receiving a series of lower and lower quotes, but stops as soon as your luck runs out (Hey 1982; Charness and Kuhn 2011).

Tit-for-tat. *Begin by cooperating, then respond in kind to your opponent; If your opponent cooperates, then cooperate; if your opponent defects, then defect* (Axelrod 1984; Rapaport, Seale, and Colman 2015).

Linear Optical Trajectory (LOT). *To intersect with another moving object, adjust your speed so that your angle of gaze remains constant.* (McBeath, Shaffer, and Kaiser 1995; Gigerenzer 2007).

Take-the-best. *To decide which of two alternatives has a higher value on a specific criterion, (i) first search the cues in order of their predictive validity; (ii) next, stop search when a cue is found which discriminates between the alternatives; (iii) then, choose the alternative selected by the discriminating cue. (iv) If all cues fail to discriminate between the two alternatives, then choose an alternative by chance* (Einhorn 1970; Gigerenzer and Goldstein 1996).

Recognition: *To decide which of two alternatives has a higher value on a specific criterion and one of the two alternatives is recognized, choose the alternative that is recognized* (Goldstein and Gigerenzer 2002; Davis-Stober, Dana, and Budescu 2010; Pachur, Todd, Gigerenzer, Schooler, and Goldstein 2012).

Fluency: *To decide which of two alternatives has a higher value on a specific criterion, if both alternatives are recognized but one is recognized faster, choose the alternative that is recognized faster*(Schooler and Hertwig 2005; Herzog and Hertwig 2013).

$\frac{1}{N}$ **Rule:** *For N feasible options, invest resources equally across all N options* (Hertwig, Davis, and Sulloway 2002; DeMiguel, Garlappi, and Uppal 2009).

There are three lines of responses to the Fast and Frugal program to mention. Take-the-Best is an example of a *non-compensatory* decision rule, which means that the first discriminating cue cannot be "compensated" by the cue-information remaining down the order. This condition, when it holds, is thought to warrant taking a decision on the first discriminating cue and ignoring the remaining cue-information. The computational efficiency of Take-the-Best is supposed to come from only evaluating a few cues, which number less than 3 on average in benchmarks tests (Czerlinski, Gigerenzer, and Goldstein 1999). However, all of the cue validities need to be known by the decision-maker and sorted before initiating the search. So, Take-the-Best by design treats a portion of the necessary computational costs to execute the heuristic as exogenous. Although the lower-bound for sorting cues by comparison is $O(n \log n)$, there is little evidence to suggest that humans sort cues by the most efficient sorting algorithms in this class. On the contrary, such operations are precisely of the kind that qualitative probability judgements demand (Section 1.2). Furthermore, in addition to the costs of ranking cue validities, there is the cost of acquisition and the determination that the agent's estimates are non-compensatory. Although the exact accounting of the cognitive effort presupposed is unknown, and argued to be lower than critics suggest (Katsikopoulos, Schooler, and Hertwig 2010), nevertheless these necessary steps threaten to render Take-the-Best non-compensatory in execution but not in what is necessary prior to setting up the model to execute.

A second line of criticism concerns the cognitive plausibility of Take the Best (Chater, Oaksford, Nakisa, and Redington 2003). Nearly all of the empirical data on the performance characteristics of Take-the-Best are by computer simulations, and those original competitions pitted Take the Best against standard statistical models (Czerlinski, Gigerenzer, and Goldstein 1999) but omitted standard machine learning algorithms that Chater, Oaksford and colleagues found performed just as well as Take the Best. Since these initial studies, the focus has shifted to machine learning, and includes variants of Take-the-Best, such as "greedy cue permutation" that performs provably better than the original and is guaranteed to always find accurate solutions when they exist (Schmitt and Martignon 2006). Setting aside criticisms targeting the comparative performance advantages of Take the Best qua decision model, others have questioned the plausibility of using Take-the-Best as a cognitive model. For example, Take-the-Best presumes that cue-information is processed serially, but the speed advantages of the model translate to an advantage in human decision-making derives only if humans process cue information on a serial architecture. If instead people process cue information on a parallel cognitive architecture, then the comparative speed advantages of Take-the-Best would become moot (Chater, Oaksford, Nakisa, and Redington 2003).

The third line of objection concerns whether the Fast-and-Frugal program truly mounts a challenge to the normative standards of optimization, dominance-reasoning, and consistency, as advertised. Take-the-Best is an algorithm for decision-making that does not comport with the axioms of expected utility theory. For one thing, its lexicographic structure violates the Archimedean axiom (Section 1.2, A2). For another, it is presumed to violate the transitivity condition of the Ordering axiom (A1). Further still, the "less-is-more" effects appear to violate Good's principle (Good 1967), a central pillar of Bayesian decision theory, which recommends to delay making a terminal decision between alternative options if the opportunity arises to acquire free information. In other words, according canonical Bayesianism, free advice is a bore but no one ought to turn down free information (Pedersen and Wheeler 2014). If noncompensatory decision rules like Take-the-Best violate Good's principle, then perhaps the whole Bayesian machinery ought to go (Gigerenzer and Brighton 2009).

But these points merely tell us that attempts to formulate Take-the-Best in terms of an ordering of prospects on a real-valued index won't do, not that ordering and numerical indices have all got to go. As we saw in Section 1.1, there is a long and sizable literature on lexicographic probabilities and non-standard analysis, including early work specifically addressing non-compensatory nonlinear models (Einhorn 1970). Second, Gigerenzer argues that "cognitive algorithms... need to meet more important constraints than internal consistency" (Gigerenzer and Goldstein 1996), which includes transitivity, and elsewhere advocates abandoning coherence as a normative standard (Arkes, Gigerenzer, and Hertwig 2016). However, Take-the-Best presupposes that cues are ordered by cue validity, which naturally entails

transitivity, otherwise Take-The-Best could neither be coherently specified nor effectively executed. More generally, the Fast and Frugal school's commitment to formulating heuristics algorithmically and implementing them as computational models commits them to the normative standards of optimization, dominance reasoning, and logical consistency.

Finally, Good's principle states that a decision-maker facing a single-person decision-problem cannot be worse (in expectation) from receiving free information. Exceptions are known in game theory (Osborne 2003, p. 283), however, that involve asymmetric information among two or more decision-makers. But there is also an exception for single-person decision-problems involving indeterminate or imprecise probabilities (Pedersen and Wheeler 2015). The point is that Good's principle is not a fundamental principle of probabilistic methods, but instead is a specific result that holds for the canonical theory of single-person decision-making with determinate probabilities.

8 Appraising Human Rationality

The rules of logic, the axioms of probability, the principles of utility theory—humans flout them all, and do so as a matter of course. But are we irrational to do so? That depends on what being rational amounts to. For a Bayesian, any qualitative comparative judgment that does not abide by the axioms of probability is, by definition, irrational. For a baker, any recipe for bread that is equal parts salt and flour is irrational, even if coherent. Yet Bayesians do not war with bakers. Why? Because bakers are satisfied with the term 'inedible' and do not aspire to commandeer 'irrational'.

The two schools of heuristics (Section 7) reach sharply different conclusions about human rationality. Yet, unlike bakers, their disagreement involves the meaning of 'rationality' and how we ought to appraise human judgment and decision making. The "rationality wars" are not the result of "rhetorical flourishes" concealing a broad consensus (Samuels, Stich, and Bishop 2002), but substantive disagreements (Section 7.2) that are obscured by ambiguous use of terms like 'rationality'.

In this section we first distinguish seven different notions of rationality, highlighting the differences in aim, scope, standards of assessment, and differences in the objects of evaluation. We then turn to consider different two importantly different normative standards used in bounded rationality, followed by an example, the *perception-cognition gap*, illustrating how slight variations of classical experimental designs in the biases and heuristics literature change both the results and the normative standards used to evaluate those results.

8.1 Rationality

While Aristotle is credited with saying that humans are rational, Bertrand Russell later confessed to searching a lifetime in vain for evidence in Aristotle's favor. Yet 'rationality' is what Marvin Minsky called a suitcase word, a term that needs to be unpacked *before* getting anywhere.

One meaning, central to decision theory, is *coherence*, which is merely the requirement that your commitments not be self-defeating. The subjective Bayesian representation of rational preference over options as inequalities in subjective expected utility delivers coherence by applying a dominance principle to (suitably structured) preferences. A closely related application of dominance reasoning is the minimization of expected loss (or maximization of expected gain in economics) according to a suitable loss function, which may even be asymmetric (Elliott, Komunjer, and Timmermann 2005) or applied to radically restricted agents, such as finite automata (Rubinstein 1986). Coherence and dominance reasoning underpin expected utility theory (Section 1.1), too.

A second meaning of rationality refers to an interpretive stance or disposition that we take to understand the beliefs, desires, and actions of another person (Dennett 1971) or to understand anything they might say in a shared language (Davidson 1974). On this view, rationality refers to a bundle of assumptions we grant to another person in order to understand their behavior, including speech. When we offer a reason-giving explanation for another person's behavior, we take such a stance. If I say "*the driver laughed because she made a joke*" you would not get far in understanding me without granting to me, and even this imaginary driver and woman, a lot. So, in contrast to the lofty normative standards

of coherence that few if any mortals meet, the standards of rationality associated with an interpretive stance are met by practically everyone.

A third meaning of rationality, due to Hume (1738), applies to your beliefs, appraising them in how well they are calibrated with your experience. If in your experience the existence of one thing is invariably followed by an experience of another, then believing that the latter follows the former is rational. We might even go so far as to say that your expectation of the latter given your experience of the former is rational. This view of rationality is an evaluation of a person's commitments, like coherence standards; but unlike coherence, Hume's notion of rationality seeks to tie the rational standing of a belief directly to evidence from the world. Much of contemporary epistemology endorses this concept of rationality while attempting to specify the conditions under which we can correctly attribute knowledge to someone's beliefs.

A fourth meaning of rationality, called *substantive rationality* by Max Weber (Weber 1905), applies to the evaluation of your aims of inquiry. Substantive rationality invokes a Kantian distinction between the worthiness of a goal, on the one hand, and how well you perform instrumentally in achieving that goal, on the other. Aiming to count the blades of grass in your lawn is arguably not a rational end to pursue, even if you were to use the instruments of rationality flawlessly to arrive at the correct count.

A fifth meaning of rationality, due to Peirce (1955) and taken up by the American pragmatists, applies the process of changing a belief rather than the Humean appraisal of a currently held belief. On Peirce's view, people are plagued by doubt not by belief; we don't expend effort testing the sturdiness of our beliefs, but rather focus on those that come into doubt. Since inquiry is pursued to remove the doubts we have, not certify the stable beliefs we already possess, principles of rationality ought to apply to the methods for removing doubt (Dewey 1960). On this view, questions of what is or is not substantively rational will be answered by the inquirer: for an agronomist interested in grass cover sufficient to crowd out an invasive weed, obtaining the grass-blade count of a lawn would be a substantively rational aim to pursue.

A sixth meaning of rationality appeals to an organism's capacities to assimilate and exploit complex information and revise or modify it when it is no longer suited to task. The object of rationality according to this notion is *effective behavior*. Jonathan Bennett discusses this notion of rationality in his case study of bees:

> All our *prima facie* cases of rationality or intelligence were based on the observation that some creature's behaviour was in certain dependable ways successful or appropriate or apt, relative to its presumed wants or needs. ...There are canons of appropriateness whereby we can ask whether an apian act is appropriate not to that which is particular and present to the bee but rather to that which is particular and past or to that which is not particular at all but universal (Bennett 1964, p. 85).

Like Hume's conception, Bennetti's view ties rationality to successful interactions with the world. Further, like the pragmatists, Bennett includes for appraisal the dynamic process rather than simply the synchronic state of one's commitments or the current merits of a goal. But unlike the pragmatists, Bennett conceives of rationality to apply to a wider range of behavior than the logic of deliberation, inquiry, and belief change.

A seventh meaning of rationality resembles the notion of coherence by defining rationality as the absence of a defect. For Bayesians, sure-loss is the epitome of *irrationality* and coherence is simply its absence. Sorensen has suggested a generalization of this strategy, one where rationality is conceived as the absence of irrationality *tout court*, just as cleanliness is the absence of dirt. Yet, owing to the long and varied ways that irrationality can arise, a consequence of this view is that there then would be no unified notion of rationality to capture the idea of thinking as one ought to think (Sorensen 1991).

These seven accounts of rationality are neither exhaustive nor complete. But they suffice to illustrate the range of differences among rationality concepts, from the objects of evaluation and the standards of assessment, to the roles, if any at all, that rationality is conceived to play in reasoning, planning, deliberation, explanation, prediction, signaling, and interpretation. One consequence of this hodgepodge

of rationality concepts is a pliancy in the attribution of irrationality that resembles Victorian methods for diagnosing the vapors. The time may have come to retire talk of rationality altogether, or to demand a specification of the objects of evaluation, the normative standards to be used for assessment, and require ample attention to the implications that follow from those commitments.

8.2 Normative Standards in Bounded Rationality

What are the standards against which our judgments and decisions ought to be evaluated? A property like systematic bias may be viewed as a fault or an advantage depending on how outcomes are scored (Sections 4). A full reckoning of the costs of operating a decision procedure may tip the balance in favor of a model that is sub-optimal when costs are no constraint, even when there is agreement of how outcomes are to be scored (Sections 2.1). Desirable behavior, such as prosocial norms, may be impossible within an idealized model but commonplace in several different types of non-idealized models (Section 5.2).

Accounts of bounded rationality typically invoke one of two types of normative standards, a coherence standard or an accuracy standard. Among the most important insights from the study of boundedly rational judgment and decision making is that, not only is it possible to meet one standard without meeting the other, but meeting one standard may inhibit meeting the other.

Coherence standards in bounded rationality typically appeal to probability, statistical decision theory, or propositional logic. The "standard picture" of rational reasoning, according to Edward Stein,

> is to reason in accordance with principles of reasoning that are based on rules of logic, probability theory, and so forth. If the standard picture of reasoning is right, principles of reasoning that are based on such rules are *normative principles of reasoning*, namely they are principles we *ought* to reason in accordance with (Stein 1996, §1.2).

Logic and probability coherence standards are usually invoked when there are experimental results pointing to violations of those standards, particularly in the heuristics and biases literature (Section 7.1). However, little is said about how and when our reasoning ought to be in accordance with these standards or even what, precisely, the normative standards of logic and probability amount to. Stein discusses the logical rule of *And-Elimination* and a normative principle for belief that it supports, one where believing the conjunction *birds sing and bees waggle* commits you rationally to believing each conjunct. Yet Stein switches to probability to discuss what principle ought to govern conjoining two beliefs. Why?

Propositional logic and probability are very different formalism (Haenni, Romeijn, Wheeler, and Williamson 2011). For one thing, the truth-functional semantics of logic is compositional whereas probability is not compositional, except when events are probabilistically independent. Why then is the elimination rule from logic and the introduction rule from probability the standard rather than the elimination rule from probability (marginalization) and the introduction rule from logic (adjunction)? Answering this question requires a positive account of what "based on", "anchored in", or other metaphorical relationships amount to. By way of comparison, there is typically no analog to the representation theorems of expected utility theory (Section 1.1) specifying the relationship between qualitative judgment and quantitative representation, and no accounting for the conditions under which that relationship holds.

The second type of normative standard assesses the accuracy of a judgment or decision making process, where the focus is getting the correct answer. Consider the accuracy of a categorical judgment, such as predicting whether a credit-card transaction is fraudulent ($Y = 1$) or legitimate ($Y = 0$). *Classification accuracy* is the number of correct predictions from all predictions made, which is often expressed as a ratio. But classification accuracy can yield a misleading assessment. For example, a method that always reported transactions as legitimate, $Y = 0$, would in fact yield a very high accuracy score ($> 97\%$) due to the very low rate ($< 3\%$) of fraudulent credit card transactions. The problem here is that classification accuracy is a poor metric for problems that involve imbalanced classes with few positive instances (i.e., few cases where $Y = 1$). More generally, a model with no predictive power can have high accuracy, and a model with comparatively lower accuracy can have greater predictive power. This observation is referred to as the *accuracy paradox*.

The accuracy paradox is one motivation for introducing other measures of predictive performance. For our fraud detection problem there are two ways your prediction can be correct and two ways it can be wrong. A prediction can be correct by predicting that $Y = 1$ when in fact a transaction is fraudulent (a *true positive*) or predicting $Y = 0$ when in fact a transaction is legitimate (a *true negative*). Correspondingly, one may err by either predicting $Y = 1$ when in fact $Y = 0$ (a *false positive*) or predicting $Y = 0$ when in fact a transaction is legitimate (a *true negative*). These four possibilities are presented in the following two-by-two contingency table, which is sometimes referred to as a *confusion matrix*:

		Actual Class	
	Y	1	0
Predicted Class	1	true positive	false positive
	0	false negative	true negative

For a binary classification problem involving *N* examples, each prediction will fall into one of these four categories. The performance of your classifier with respect to those *N* examples can then be assessed. A perfectly inaccurate classifier will have all zeros in the diagonal; a perfectly accurate classifier will have all zeros in the counterdiagonal. The *precision* of your classifer is the ratio of true positives to all positive predictions, that is *true positives* / (*true positives* + *false positives*). The *recall* of your classifier is the ratio of true positives to all true predictions, that is *true positives* / (*true positives* + *false negatives*).

There are two points to notice. The first is that in practice there is typically a trade-off between precision and recall, and the costs to you of each will vary from one problem to another. A trade-off of precision and recall that suits detecting credit card fraud may not suit detecting cancer, even if the frequencies of positive instances are identical. The point of training a classifier on known data is to make predictions on out of sample instances. So, tuning your classifier to yield a suitable trade-off between precision and recall in your training data is no guarantee that you will see this trade-off generalize.

The moral is that to evaluate the performance of your classifier it is necessary to specify the purpose for making the classification and even then good performance on your training data may not generalize. None of this is antithetical to coherence reasoning per se, as we are making comparative judgments and reasoning by dominance. But putting the argument in terms of coherence changes the *objects* of evaluation, moving from the point of view from the first person the decision maker to that of a third person decision modeler.

8.3 The Perception-Cognition Gap

Do human beings systematically violate the norms of probability and statistics? Petersen and Beach (1967) thought not. On their view human beings are intuitive statisticians (Section 5.1), so probability theory and statistics are a good, first approximation of human judgment and decision making. Yet, just as their optimistic review appeared to cement a consensus view about human rationality, Amos Tversky and Daniel Kahneman began their work to undo it. People are particularly bad at probability and statistics, the heuristics and biases program (Section 7.1) found, so probability theory, statistics, and even logic do not offer a good approximation of human decision making. One controversy over these negative findings concerns the causes of those effects—whether the observed responses point to minor flaws in otherwise adaptive human behavior or something much less charitable about our habits and constitution.

In contrast to this poor showing on cognitive tasks, people are generally thought to be optimal or near-optimal in performing low-level motor control and perception tasks. Planning goal-directed movement, like pressing an elevator button with your finger or placing your foot on a slippery river stone, requires your motor control system to pick one among a dizzying number of possible movement strategies to achieve your goal while minimizing biomechanical costs (Trommershäuser, Maloney, and Landy 2003). The loss function that our motor control system appears to use increases approximately quadratically with error for small errors but significantly less for large errors, suggesting that our motor control

system is also robust to outliers (Körding and Wolpert 2004). What is more, advances in machine learning have been guided by treating human performance errors for a range of perception tasks as proxies for Bayes error, yielding an observable, near-perfect normative standard. Unlike cognitive decisions, there is very little controversy concerning the overall optimality of our motor-perceptual decisions. This difference between high-level and low-level decisions is called the *perception-cognition gap*.

Some view the perception-cognition gap as evidence for the claim that people use fundamentally different strategies for each type of task (Section 7.2). An approximation of an optimal method is not necessarily an optimal approximation of that method, and the study of cognitive judgments and deliberative decision-making is led astray by assuming otherwise (Mongin 2000). Another view of the perception-cognition gap is that it is largely an artifact of methodological differences across studies rather than a robust feature of human behavior. We review evidence for this second argument here.

Classical studies of decision-making present choice problems to subjects where probabilities are described. For example, you might be asked to choose the prospect of winning €300 with probability 0.25 or the prospect of winning €400 with probability 0.2. Here, subjects are given a numerical description of probabilities, are typically asked to make one-shot decisions without feedback, and their responses are found to deviate from the expected utility hypothesis. However, in motor control tasks, subjects have to use internal, implicit estimates of probabilities, often learned with feedback, and these internal estimates are near optimal. Are perceptual-motor control decisions better because they provide feedback whereas classical decision tasks do not, or are perceptual-motor control decisions better because they are non-cognitive?

Jarvstad et al. (2013) explored the robustness of the perception-cognition gap by designing (a) a finger-pointing task that involved varying target sizes on a touch-screen computer display; (b) an arithmetic learning task involving summing four numbers and accepting or rejecting a proposed answer with a target tolerance, where the tolerance range varied from problem to problem, analogous to the width of the target in the motor-control task; and (c) a standard classical probability judgment task that involved computing the expected value of two prospects. The probability information across the tasks was in three formats: low-level, high-level, and classical, respectively.

Once confounding factors across the three types of tasks are controlled for, Jarvstad et al.'s results suggest that (i) the perception-cognition gap is largely explained by differences in how performance is assessed; (ii) the *decisions by experience* vs *decisions by description* gap (Hertwig, Barron, Weber, and Erev 2004) is due to assuming that exogenous objective probabilities and subjective probabilities match; (iii) people's ability to make high-level decisions is better than the biases and heuristics literature suggests (Section 7.1); and (iv) differences between subjects are more important for predicting performance than differences between the choice tasks (Jarvstad, Hahn, Rushton, and Warren 2013).

The upshot, then, is that once the methodological differences are controlled for, the perception-cognition gap appears to be an artifact of two different normative standards applied to tasks. If the standards applied to assessing perceptual-motor tasks are applied to classical cognitive decision-making tasks, then both appear to perform well. If instead the standards used for assessing the classical cognitive tasks are applied to perceptual-motor tasks, then both will appear to perform poorly.

Acknowledgements

Thanks to Sebastian Ebert, Ulrike Hahn, Ralph Hertwig, Konstantinos Katsikopoulos, Jan Nagler, Christine Tiefensee, Conor Mayo-Wilson, and an anonymous referee for helpful comments on earlier drafts of this article.

References

Alexander, J. M. (2007). *The Structural Evolution of Morality*. New York: Cambridge University Press.

Alexander, R. D. (1987). *The Biology of Moral Systems*. London: Routledge.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de

l'école américaine. *Econometrica 21*(4), 503–546.

Anand, P. (1987). Are the preference axioms really rational? *Theory and Decision 23*, 189–214.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review 98*, 409–429.

Anderson, J. R. and L. J. Schooler (1991). Reflections of the environment in memory. *Psychological Science 2*, 396–408.

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

Arló-Costa, H. and A. P. Pedersen (2011). Bounded rationality: Models for some fast and frugal heuristics. In A. Gupta, J. van Benthem, and E. Pacuit (Eds.), *Games, Norms and Reasons: Logic at the Crossroads*. Springer.

Arrow, K. (2004). Is bounded rationality unboundedly rational? Some ruminations. In M. Augier and J. G. March (Eds.), *Models of a man: Essays in memory of Herbert A. Simon*, Cambridge, MA, pp. 47–55. MIT Press.

Aumann, R. J. (1962). Utility theory without the completeness axiom. *Econometrica 30*, 445–462.

Aumann, R. J. (1997). Rationality and bounded rationality. *Games and Economic Behavior 21*, 2–17.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Ballard, D. H. and C. M. Brown (1982). *Computer Vision*. Englewood Cliffs, NJ: Prentice Hall.

Bar Hillel, M. and A. Margalit (1988). How vicious are cycles of intransitive choice? *Theory and Decision 24*, 119–145.

Bar Hillel, M. and W. A. Wagenaar (1991). The perception of randomness. *Advances in Applied Mathematics 12*(4), 428–454.

Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *Science 286*(5439), 509–512.

Barkow, J., L. Cosmides, and J. Tooby (Eds.) (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Baumeister, R. F., E. Bratslavsky, and C. Finkenauer (2001). Bad is stronger than good. *Review of General Psychology 5*(4), 323–370.

Bazerman, M. H. and D. A. Moore (2008). *Judgment in Managerial Decision Making* (7th ed.). New York: Wiley.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research 30*(5), 961–981.

Bennett, J. (1964). *Rationality: An Essay towards an Analysis*. London: Routledge.

Berger, J. O. (1980). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer.

Bernoulli, D. (1954/1738). Exposition of a new theory on the measurement of risk. *Econometrica 22*(1), 23–36. Trans. Louise Sommer, from "Specimen Theoriae Novae de Mensura Sortis", *Commentarii Academiae Scientiarium Imperialis Petropolitanae*, Tomus V, 1738, pp. 175–192.

Bewley, T. S. (2002). Knightian decision theory: Part I. *Decisions in Economics and Finance 25*, 79–110. Reprint of *Cowes Foundation* Discussion Paper 807, 1986.

Bicchieri, C. (2005). *The Grammar of Society*. New York: Cambridge University Press.

Bicchieri, C. and R. Muldoon (2014). Social norms. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.). Metaphysics Research Lab, Stanford University.

Birnbaum, M. H. (1979). Base rates in Baysian inference: Signal detection analysis of the cab problem. *The American Journal of Psychology 96*(1), 85–94.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Blume, L., A. Brandenburger, and E. Dekel (1991). Lexicographic probabilities and choice under uncertainty. *Econometrica 58*(1), 61–78.

Bonet, B. and H. Geffner (2001). Planning as heuristic search. *Artificial Intelligence 129*(1–2), 5–33.

Bowles, S. and H. Gintis (2011). *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton, NJ: Princeton University Press.

Boyd, R. and P. J. Richerson (2005). *The Origin and Evolution of Cultures*. New York: Oxford University Press.

Brickhill, H. and L. Horsten (2016, August). Popper functions, lexicographic probability, and non-Archimedean probability. *arXiv:1608.02850v1*.

Brown, S. D. and A. Heathcote (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology 57*, 153–178.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review 50*(3), 255–272.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review 62*(3), 193–217.

Charness, G. and P. J. Kuhn (2011). Lab labor: What can labor economists learn from the lab? In *Handbook of Labor Economics*, Volume 4, pp. 229–330. Elsevier.

Chater, N. (2014). Cognitive science as an interface between rational and mechanistic explanation. *Topics in Cognitive Science 6*, 331–337.

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

Clark, A. and D. Chalmers (1998). The extended mind. *Analysis 58*(1), 7–19.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences 4*(3), 317–331.

Coletti, G. and R. Scozzafava (2002). *Probabilistic Logic in a Coherent Setting*. Trends in logic, 15. Dordrecht: Kluwer.

Collins, P. J., K. Krzyż, S. Hartmann, G. Wheeler, and U. Hahn (2018, January). Conditionals and testimony. Unpublished Manuscript.

Czerlinski, J., G. Gigerenzer, and D. G. Goldstein (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, and T. A. Group (Eds.), *Simple Heuristics that Make Us Smart*, pp. 97–118. Oxford University Press.

Damore, J. A. and J. Gore (2012). Understanding microbial cooperation. *Journal of Theoretical Biology 299*, 31–41.

Dana, J. and R. M. Dawes (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics 29*(3), 317–331.

Darwin, C. (1871). *The Descent of Man*. New York: Penguin Classics.

Davidson, D. (1974). Belief and the basis of meaning. *Synthese 27*(3–4), 309–323.

Davis-Stober, C. P., J. Dana, and D. V. Budescu (2010). Why recognition is rational: Optimality results on single-variable decision rules. *Judgment and Decision Making 5*(4), 216–229.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist 34*(7), 571–582.

de Finetti, B. (1974). *Theory of Probability: A critical introductory treatment*, Volume 1 and 2. Wiley.

de Finetti, B. and L. J. Savage (1962). Sul modo di scegliere le probabilità iniziali. *Biblioteca del Metron, Serie C 1*, 81–154.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the $\frac{1}{N}$ portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

Dennett, D. C. (1971). Intentional systems. *Journal of Philosophy 68*(4), 87–106.

Dewey, J. (1960). *The Quest for Certainty*. Gifford Lectures of 1929. New York: Capricorn Books.

Dhami, M. K., R. Hertwig, and U. Hoffrage (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin 130*(6), 959–988.

Domingos, P. (2000). A unified bias-variance decomposition and its applications. In *Proceedings of the 17th International Conference on Machine Learning*, pp. 231–238. Morgan Kaufmann.

Doyen, S., O. Klein, C.-L. Pichton, and A. Cleeremans (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS One 7*(1), e29081.

Dubins, L. E. (1975). Finitely additive conditional probability, conglomerability, and disintegrations. *Annals of Probability 3*, 89–99.

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

Elliott, G., I. Komunjer, and A. Timmermann (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies 72*, 1107–1125.

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics 75*, 643–69.

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

Fennema, H. and P. Wakker (1997). Original and cumulative prospect theory: A discussion of empirical differences. *Journal of Behavioral Decision Making 10*, 53–64.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research 50*, 123–129.

Fiedler, K. and P. Juslin (2006). *Information Sampling and Adaptive Cognition*. Cambridge: Cambridge University Press.

Fishburn, P. C. (1982). *The Foundations of Expected Utility*. Dordrecht: D. Reidel.

Fisher, R. A. (1936). Uncertain inference. *Proceedings of the American Academy of Arts and Sciences 71*, 245–258.

Forscher, P., C. K. Lai, J. R. Axt, C. R. Ebersole, M. Herman, P. G. Devine, and B. A. Nosek (2017, July). A meta-analysis of change in implicit bias. Unpublished Manuscript. Under review.

Friedman, J. (1997). On bias, variance, 0-1 loss and the curse of dimensionality. *Journal of Data Mining and Knowledge Discovery 1*, 55–77.

Friedman, M. (1953). The methodology of positive economics. In *Essays in Positive Economics*, pp. 3–43. University of Chicago Press.

Friedman, M. and L. J. Savage (1948). The utility analysis of choices involving risk. *Journal of Political Economy 56*, 279–304.

Friston, K. (2010). The free-energy principle: A unified brain theory. *Nature Reviews Neuroscience 11*, 127–138.

Galaabaatar, T. and E. Karni (2013). Subjective expected utility with incomplete preferences. *Econometrica 81*(1), 255–284.

Gergely, G., H. Bekkering, and I. Király (2002). Developmental psychology: Rational imitation in preverbal infants. *Nature 415*, 755–756.

Ghallab, M., D. Nau, and P. Traverso (2016). *Automated Planning and Acting*. New York: Cambridge University Press.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review 103*(3), 592–596.

Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious*. New York: Viking Press.

Gigerenzer, G. and H. Brighton (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science 1*(1), 107–43.

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

Gigerenzer, G., W. Hell, and H. Blank (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance 14*(3), 513–525.

Gigerenzer, G., R. Hertwig, and T. Pachur (Eds.) (2011). *Heuristics: The Foundations of Adaptive Behavior*. New York: Oxford University Press.

Gigerenzer, G., P. M. Todd, and T. A. Group (Eds.) (1999). *Simple Heuristics that Make Us Smart*. Oxford University Press.

Giles, R. (1976). A logic for subjective belief. In W. Harper and C. A. Hooker (Eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Volume I. Dordrecht: Reidel.

Giron, F. J. and S. Rios (1980). Quasi-Bayesian behavior: A more realistic approach to decision making? *Trabajos de Estadistica Y de Investigacion Operativa 31*(1), 17–38.

Glymour, C. (2001). *The Mind's Arrows*. Cambridge, MA: MIT Press.

Goldblatt, R. (1998). *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis*. Graduate Texts in Mathematics. New York: Springer-Verlag.

Goldstein, D. and G. Gigerenzer (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review 109*(1), 75–90.

Golub, B. and M. O. Jackson (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal of Microeconomics 2*(1), 112–149.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B 14*(1), 107–114.

Good, I. J. (1967). On the principle of total evidence. *The British Journal for the Philosophy of Science 17*(4), 319–321.

Good, I. J. (1983). Twenty-seven principles of rationality. In *Good Thinking: The Foundations of Probability and its Applications*, pp. 15–20. Minneapolis: University of Minnesota Press.

Güth, W., R. Schmittberger, and B. Schwarze (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization 3*(4), 367–388.

Hacking, I. (1967). Slightly more realistic personal probability. *Philosophy of Science 34*(4), 311–325.

Haenni, R., J.-W. Romeijn, G. Wheeler, and J. Williamson (2011). *Probabilistic Logics and Probabilistic Networks*. Synthese Library. Dordrecht: Springer.

Hahn, U. and P. A. Warren (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review 116*(2), 454–461. See correction in 116(4).

Halpern, J. Y. (2010). Lexicographic probability, conditional probability, and nonstandard probability. *Games and Economic Behavior 68*(1), 155–179.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review 62*(4), 255–262.

Hammond, K. R., C. J. Hursch, and F. J. Todd (1964). Analyzing the components of clinical inference. *Psychological Review 71*(6), 438–456.

Hammond, P. J. (1994). Elementary non-Archimedean representations of probability for decision theory and games. In P. Humphreys (Ed.), *Patrick Suppes: Scientific Philosopher*, Volume 1: Probability and Probabilistic Causality, pp. 25–59. Dordrecht, The Netherlands: Kluwer.

Haykin, S. O. (2013). *Adaptive Filter Theory* (5th ed.). London: Pearson.

Henrich, J. and F. J. Gil-White (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior 22*(3), 165–196.

Hertwig, R., G. Barron, E. U. Weber, and I. Erev (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science 15*(8), 534–539.

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

Hertwig, R. and G. Gigerenzer (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making 12*, 275–305.

Hertwig, R. and T. J. Pleskac (2008). The game of life: How small samples render choice simpler. In *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*, pp. 209–235. Oxford: Oxford University Press.

Herzog, S. and R. Hertwig (2013). The ecological validity of fluency. In C. Unkelbach and R. Greifeneder (Eds.), *The Experience of Thinking: How Fluency of Mental Processes Influences Cognition and Behavior*, pp. 190–219. Psychology Press.

Hey, J. D. (1982). Search for rules for search. *Journal of Economic Behavior and Organization 3*(1), 65–81.

Ho, T.-H. (1996). Finite automata play repeated prisoner's dilemma with information processing costs. *Journal of Economic Dynamics and Control 20*, 173–207.

Hochman, G. and E. Yechiam (2011). Loss aversion in the eye and in the heart. *Journal of Behavioral Decision Making 24*(2), 140–156.

Hogarth, R. M. (2012). When simple is hard to accept. In P. M. Todd, G. Gigerenzer, and T. A. Group (Eds.), *Ecological Rationality: Intelligence in the World*, pp. 61–79. New York: Oxford University Press.

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

Howe, M. L. (2011). The adaptive nature of memory and its illusions. *Current Directions in Psychological Science 20*(5), 312–315.

Hume, D. (1738). *A Treatise of Human Nature*. Version by Jonathan Bennett, 2008: www.earlymoderntexts.com.

Hutchinson, J. M., C. Fanselow, and P. M. Todd (2012). Car parking as a game between simple heuristics. In P. M. Todd, G. Gigerenzer, and T. A. Group (Eds.), *Ecological Rationality: Intelligence in the World*, pp. 454–484. New York: Oxford University Press.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jarvstad, A., U. Hahn, S. K. Rushton, and P. A. Warren (2013). Perceptuo-motor, cognitive, and description-based decision-making seem equally good. *Proceedings of the National Academy of Sciences 110*(40), 16271–16276.

Jevons, W. S. (1871). *The Theory of Political Economy*. Palgrave Classics in Economics. London: Macmillian and Company.

Juslin, P. and H. Olsson (2005). Capacity limitations and the detection of correlations: Comment on Kareev. *Psychological Review 112*(1), 256–267.

Juslin, P., A. Winman, and P. Hannson (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervalsve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review 114*(3), 678–703.

Kahneman, D. (2017). Reply to Schimmack, Heene, and Kesavan's 'Rconstruction of a Train Wreck: How Priming Research Went Off the Rails'. https://replicationindex.wordpress.com/2017/02/02/reconstruction-of-a-train-wreck-how-priming-research-went-of-the-rails/comment-page-1/#comment-1454.

Kahneman, D., B. Slovic, and A. Tversky (Eds.) (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology 3*, 430–454.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*, 263–291.

Kahneman, D. and A. Tversky (1996). On the reality of cognitive illusions. *Psychological Review 103*(3), 582–591.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition 56*(3), 263–269.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review 107*(2), 397–402.

Karni, E. (1985). *Decision Making Under Uncertainty: The Case of State-Dependent Preferences*. Cambridge, MA: Harvard University.

Katsikopoulos, K. V. (2010). The less-is-more effect: Predictions and tests. *Judgment and Decision Making 5*(4), 244–257.

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

Kaufmann, E. and W. W. Wittmann (2016). The success of linear bootstrapping models: Decision domain-, expertise-, and criterion-specific meta-analysis. *PLoS One 11*(6), e0157914.

Keeney, R. L. and H. Raiffa (1976). *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. New York: Wiley.

Kelly, K. T. and O. Schulte (1995). The computable testability of theories making uncomputable predictions. *Erkenntnis 43*(1), 29–66.

Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.

Kidd, C. and B. Y. Hayden (2015). The psychology and neuroscience of curiosity. *Neuron 88*(3), 449–460.

Kirsch, D. (1995). The intelligence use of space. *Artificial Intelligence 73*(1–2), 31–68.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences 19*, 1–53.

Koopman, B. O. (1940). The axioms and algebra of intuitive probability. *Annals of Mathematics 41*(2), 269–292.

Körding, K. P. and D. M. Wolpert (2004). The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences 101*, 9839–42.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory 27*(2), 245–252.

Kühberger, A., M. Schulte-Mecklenbeck, and J. Perner (1999). The effects of framing, reflection, probability, and payoff on risk preference in choice tasks. *Organizational Behavior and Human Decision Processes 78*(3), 204–231.

Kyburg, Jr., H. E. (1978). Subjective probability: Criticisms, reflections, and problems. *Journal of Philosophical Logic 7*(1), 157–180.

Levi, I. (1977). Direct inference. *Journal of Philosophy 74*, 5–29.

Levi, I. (1983). Who commits the base-rate fallacy? *Behavioral and Brain Sciences 6*(3), 502–506.

Lewis, R. L., A. Howes, and S. Singh (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science 6*, 279–311.

Lichtenberg, J. M. and Özgür Simsek (2016). Simple regression models. *Proceedings of Machine Learning Research 58*, 13–25.

Loomes, G. and R. Sugden (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal 92*(4), 805–824.

Loridan, P. (1984). ε-solutions in vector minimization problems. *Journal of Optimization Theory and Applications 43*(2), 265–276.

Luce, R. D. and H. Raiffa (1957). *Games and Decisions: Introduction and Critical Suvey*. New York: Dover.

Marr, D. C. (1982). *Vision*. New York: Freeman.

May, K. O. (1954). Intransitivity, utility, and the aggregation of preference patterns. *Econometrica 22*(1), 1–13.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

McBeath, M. K., D. M. Shaffer, and M. K. Kaiser (1995). How baseball outfielders determine where to run to catch fly balls. *Science 268*(5210), 569–573.

McNamara, J. M., P. C. Trimmer, and A. I. Houston (2014). Natural selection can favour 'irrational' behavior. *Biology Letters 10*(1), 20130935.

Meder, B., R. Mayrhofer, and M. Waldmann (2014). Structural induction in diagnostic causal reasoning. *Psychological Review 121*(3), 277–301.

Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: Minnesota Press.

Mill, J. S. (1844). On the definition of political economy. In J. M. Robson (Ed.), *The Collected Works of John Stuart Mill*, Volume IV of *Essays on Economics and Society, Part I*. Toronto: University of Toronto Press.

Mongin, P. (2000). Does optimization imply rationality. *Synthese 124*(1–2), 73–111.

Nau, R. (2006). The shape of incomplete preferences. *The Annals of Statistics 34*(5), 2430–2448.

Newell, A. and H. A. Simon (1956, June). The logic theory machine: A complex information processing system. Technical Report P-868, The Rand Corporation, Santa Monica, CA.

Newell, A. and H. A. Simon (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newell, A. and H. A. Simon (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM 19*(3), 113–126.

Neyman, A. (1985). Bounded complexity justifies cooperation in the finitely repeated prisoner's dilemma. *Economic Letters 19*(3), 227–229.

Norton, M. I., D. Mochon, and D. Ariely (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology 22*(3), 453–460.

Nowak, M. A. and R. M. May (1992). Evolutionary games and spatial chaos. *Nature 359*, 826–829.

Oaksford, M. and N. Chater (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review 101*(4), 608–631.

Oaksford, M. and N. Chater (2007). *Bayesian Rationality*. Oxford: Oxford University Press.

Ok, E. A. (2002). Utility representation of an incomplete preference relation. *Journal of Economic Theory 104*(2), 429–449.

Osborne, M. J. (2003). *An Introduction to Game Theory*. Oxford: Oxford University Press.

Oswald, F. L., G. Mitchell, H. Blanton, J. Jaccard, and P. E. Tellock (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personal Psychology 105*(2), 171–192.

Pachur, T., P. M. Todd, G. Gigerenzer, L. J. Schooler, and D. Goldstein (2012). When is the recognition heuristic an adaptive tool? In P. M. Todd, G. Gigerenzer, and T. A. Group (Eds.), *Ecological Rationality: Intelligence in the World*, pp. 113–143. New York: Oxford University Press.

Palmer, S. E. (1999). *Vision Science*. Cambridge, MA: MIT Press.

Papadimitriou, C. H. and M. Yannakakis (1994). On complexity as bounded rationality. In *Proceedings of the 26th annual ACM Symposium on Theory of Computing*, Montreal, Quebec, pp. 726–733.

Parikh, R. (1971). Existence and feasibility in arithmetic. *Journal of Symbolic Logic 36*(3), 494–508.

Payne, J. W., J. R. Bettman, and E. J. Johnson (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition 14*(3), 534–552.

Pedersen, A. P. (2014). Comparative expectations. *Studia Logica 102*(4), 811–848.

Pedersen, A. P. and G. Wheeler (2014). Demystifying dilation. *Erkenntnis 79*(6), 1305–1342.

Pedersen, A. P. and G. Wheeler (2015). Dilation, disintegrations, and delayed decisions. In *Proceedings of the 9th Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Pescara, Italy, pp. 227–236.

Peirce, C. S. (1955). *Philosophical Writings of Peirce*. New York: Dover.

Peterson, C. R. and L. R. Beach (1967). Man as an intuitive statistician. *Psychological Bulletin 68*(1), 29–46.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. London: Routledge.

Puranam, P., N. Stieglitz, M. Osman, and M. M. Pillutla (2015). Modelling bounded rationality in organizations: Progress and prospects. *The Acadamy of Management Annals 9*(1), 337–392.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization 3*, 323–343.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica 68*(5), 1281–1292.

Rapaport, A., D. A. Seale, and A. M. Colman (2015). Is Tit-for-Tat the answer? On the conclusions drawn from Axelrod's tournaments. *PLoS One 10*(7), e0134128.

Rapoport, A. and A. Chammah (1965). *Prisoner's Dilemma: A study in conflict and cooperation*. Ann Arbor: University of Michigan Press.

Regenwetter, M., J. Dana, and C. P. Davis-Stober (2011). Transitivity of preferences. *Psychological Review 118*(1), 42–56.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence 13*, 81–132.

Renyi, A. (1955). On a new axiomatic theory of probability. *Acta Math. Acad. Sci. Hungarian 6*, 285–335.

Rick, S. (2011). Losses, gains, and brains: Neuroeconomics can help to answer open questions about loss aversion. *Journal of Consumer Psychology 21*, 453–463.

Rieskamp, J. and A. Dieckmann (2012). Redundancy: Environment structure that simple heuristics can exploit. In P. M. Todd, G. Gigerenzer, and T. A. Group (Eds.), *Ecological Rationality: Intelligence in the World*, pp. 187–215. New York: Oxford University Press.

Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. *Journal of Economic Theory 39*(1), 83–96.

Russell, S. J. and D. Subramanian (1995). Provably bounded-optimal agents. *Journal of Artificial Intelligence Research 2*, 575–609.

Samuels, R., S. Stich, and M. Bishop (2002). Ending the rationality wars: How to make disputes about human rationality disappear. In R. Elio (Ed.), *Common Sense, Reasoning, and Rationality*. New York: Oxford University Press.

Samuelson, P. (1947). *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press.

Santos, F. C., M. D. Santos, and J. M. Pacheco (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature 454*, 213–2016.

Savage, L. J. (1954). *Foundations of Statistics*. New York: Wiley.

Savage, L. J. (1967, April). Difficulties in the theory of personal probability. *Philosophy of Science 34*(4), 311–325.

Schervish, M. J., T. Seidenfeld, and J. B. Kadane (2012). Measures of incoherence: How not to gamble if you must, with discussion. In J. Bernardo, A. P. Dawid, J. O. Berger, M. West, D. Heckerman, M. Bayarri, and

A. F. M. Smith (Eds.), *Bayesian Statistics 7: Proceedings of the 7th Valencia International Meeting*, Oxford Science Publications, Oxford, pp. 385–402. Clarendon Press.

Schick, F. (1986). Dutch bookies and money pumps. *Journal of Philosophy 83*(2), 112–119.

Schmitt, M. and L. Martignon (2006). On the complexity of learning lexicographic strategies. *Journal of Machine Learning Research 7*, 55–83.

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

Seidenfeld, T., M. J. Schervish, and J. B. Kadane (1995). A representation of partially ordered preferences. *The Annals of Statistics 23*, 2168–2217.

Seidenfeld, T., M. J. Schervish, and J. B. Kadane (2012). What kind of uncertainty is that? Using personal probability for expressing one's thinking about logical and mathematical propositions. *Journal of Philosophy 109*(8-9), 516–533.

Selten, R. (1998). Aspiration adoption theory. *Journal of Mathematical Psychology 42*(2–3), 191–214.

Simon, H. A. (1947). *Administrative Behavior: a study of decision-making processes in administrative organization* (1st ed.). New York: Macmillan.

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

Simon, H. A. (1955b). On a class of skew distribution functions. *Biometrika 42*(3–4), 425–440.

Simon, H. A. (1957). *Administrative Behavior: a study of decision-making processes in administrative organization* (2nd ed.). New York: Macmillan.

Simon, H. A. (1976). From substantive to procedural rationality. In T. Kastelein, S. Kuipers, W. Nijenhuis, and G. Wagenaar (Eds.), *25 Years of Economic Theory*, pp. 65–86. Boston: Springer.

Skyrms, B. (2003). *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.

Sorensen, R. A. (1991). Rationality as an absolute concept. *Philosophy 66*(258), 473–486.

Spirtes, P. (2010). Introduction to causal inference. *Journal of Machine Learning Research 11*, 1643–1662.

Stalnaker, R. (1991). The problem of logical omniscience, I. *Synthese 89*(3), 425–440.

Stanovich, K. E. and R. F. West (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences 23*(5), 645–65.

Stein, E. (1996). *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*. Oxford: Clarendon Press.

Stevens, J. R., J. Volstorf, L. J. Schooler, and J. Rieskamp (2011). Forgetting constrains the emergence of cooperative decision strategies. *Frontiers in Psychology 1*(235), 1–12.

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

Tarski, A., A. Mostowski, and R. M. Robinson (1953). *Undecidable Theories*. North-Holland Publishing Co.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization 1*(1), 39–60.

Thaler, R. H. and C. R. Sustein (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Todd, P. M., G. Gigerenzer, and T. A. Group (Eds.) (2012). *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press.

Todd, P. M. and G. F. Miller (1999). From pride and prejudice to persuasion: Satisficing in mate search. In G. Gigerenzer, P. M. Todd, and T. A. Group (Eds.), *Simple Heuristics that Make Us Smart*, pp. 287–308. Oxford University Press.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology 46*(1), 35–57.

Trommershäuser, J., L. T. Maloney, and M. S. Landy (2003). Statistical decision theory and trade-offs in the control of motor response. *Spatial Vision 16*(3), 255–275.

Turner, B. M., C. A. Rodriguez, T. M. Norcia, S. M. McClure, and M. Steyvers (2016). Why more is better: Simultaneous modeling of EEG, fMRI, and behavioral data. *NeuroImage 128*, 96–115.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review 76*, 31–48.

Tversky, A. and D. Kahneman (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology 5*(2), 207–232.

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

Tversky, A. and D. Kahneman (1977, October). Causal schemata in judgments under uncertainty. Technical Report TR-1060-77-10, Defense Advanced Research Projects Agency (DARPA).

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science 211*(4481), 483–458.

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

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

von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Vranas, P. B. (2000). Gigerenzer's normative critique of Kahneman and Tversky. *Cognition 76*, 179–193.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge: Cambridge University Press.

Waldmann, M. R., K. J. Holyoak, and A. Fratianne (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General 124*(2), 181–206.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

Weber, M. (1905). *The Protestant Ethic and the Spirit of Capitalism*. London: Allen and Unwin. Translated by Talcott Parsons (1930).

Wheeler, G. (2004). A resource bounded default logic. In J. Delgrande and T. Schaub (Eds.), *10th International Workshop on Non-Monotonic Reasoning (NMR 2004)*, Whistler, Canada, pp. 416–422.

Wheeler, G. (2017). Machine epistemology and big data. In L. McIntyre and A. Rosenberg (Eds.), *The Routledge Companion to Philosophy of Social Science*, pp. 321–329. Routledge.

Wheeler, G. and F. G. Cozman (2018, September). On the imprecision of full conditional probabilities. Unpublished Manuscript.

White, D. J. (1986). Epsilon efficiency. *Journal of Optimization Theory and Applications 49*(2), 319–337.

Yechiam, E. and G. Hochman (2014). Loss attention in a dual task setting. *Psychological Science 25*(2), 294–502.

Yule, G. U. (1911). A mathematical theory of evolution, based on the conclusions of dr. j. c. williss, f.r.s. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character 213*, 21–87.

Zaffalon, M. and E. Miranda (2015). Desirability and the birth of incomplete preferences. *ArXiv e-prints*, abs/1506.00529.