TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 11 (2019) 7–36

ISSN: 1756-8765 online
DOI: 10.1111/tops.12410

This article is part of the topic "Visions of Cognitive Science (Topic continuation)," Wayne D. Gray (Topic Editor). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview

# Cognitive Success: A Consequentialist Account of Rationality in Cognition

Gerhard Schurz,[a] Ralph Hertwig[b]

[a]*Düsseldorf Center for Logic and Philosophy of Science, Heinrich Heine University*
[b]*Center for Adaptive Rationality, Max Planck Institute for Human Development*

Received 30 October 2018; received in revised form 19 November 2018; accepted 30 November 2018

**Abstract**

One of the most discussed issues in psychology—presently and in the past—is how to define and measure the extent to which human cognition is rational. The rationality of human cognition is often evaluated in terms of normative standards based on a priori intuitions. Yet this approach has been challenged by two recent developments in psychology that we review in this article: ecological rationality and descriptivism. Going beyond these contributions, we consider it a good moment for psychologists and philosophers to join forces and work toward a new foundation for the definition of rational cognition. We take a first step in this direction by proposing that the rationality of both cognitive and normative systems can be measured in terms of their cognitive success. Cognitive success can be defined and gauged in terms of two factors: ecological validity (the system's validity in conditions in which it is applicable) and the system's applicability (the scope of conditions under which it can be applied). As we show, prominent systems of reasoning—deductive reasoning, Bayesian reasoning, uncertain conditionals, and prediction and choice—perform rather differently on these two factors. Furthermore, we demonstrate that conceptualizing rationality according to its cognitive success offers a new perspective on the time-honored relationship between the descriptive ("is") and the normative ("ought") in psychology and philosophy.

*Keywords:* Rationality; Logic; Cognitive success; Apriorism; Consequentialism; Is−ought relation

Correspondence should be sent to Ralph Hertwig, Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: hertwig@mpib-berlin.mpg.de

## 1. How psychologists measure rational cognition

For a number of decades, psychologists have typically employed only one experimental method to study whether human cognition is rational (Lopes, 1991). Their approach—devising two or more alternative hypotheses and a crucial experiment with alternative possible outcomes, each excluding one or more of the hypotheses—has been interpreted as enabling a *strong inference* (Platt, 1964). In research on the rationality of human cognition this means that the experimental set-up has been designed such that the data, people's cognitive behavior (reasoning, inference, judgment, or choice), support one of two possible results: Either individuals behave in accord with the chosen benchmark of rationality, or their cognitive behavior, measured against the benchmark, is irrational (and sometimes either deviation from the benchmark has been treated as a sign of irrationality as, for instance, in the case of the conjectures that people *neglect* base rates or *pay too much attention* to them or that people suffer from both the gambler's fallacy and the hot-hand fallacy; see Hertwig & Todd, 2000). Crucially, the benchmarks against which these evaluations are made—and human cognition is found to be rational or not—are commonly assumed to be incontrovertible. That is, the benchmarks are understood to be relatively universal, purpose invariant, content free, and domain general. Their claim to legitimacy often rests on "a priori intuitions"—a notion to which we return later. One of these seemingly incontrovertible benchmarks is the canon of classical logic. Take, for illustration, Wason's influential work on human reasoning (e.g., Wason, 1959, 1960). Far from mincing their words, Wason and Johnson–Laird argued that

> a fallacious inference, in fact, is in some ways like both an optical illusion and a pathological delusion. . . . and like most pathological delusions, we have encountered cases in which the subjects seem to reveal a stubborn resistance to enlightenment. (Wason & Johnson-Laird, 1972, p. 6)

This dooming verdict is especially notable because long before Wason, other psychologists concerned with the investigation of reasoning processes strongly opposed the use of logic to define rational thought. An example is Wilhelm Wundt, who equally unequivocally argued that

> at first it was thought that the surest way would be to take as a foundation for the psychological analysis of the thought-processes the laws of logical thinking, as they had been laid down from the time of Aristotle. . . . These norms . . . only apply to a small part of the thought-processes. Any attempt to explain, out of these norms, thought . . . can only lead to an entanglement of the real facts in a net of logical reflections. (1912/1973, pp. 148–149)

Wundt doubted that classical logic could serve as the bedrock for descriptive theories of reasoning beyond a "small part" of cognition. By extension, he rejected logic's normative claim for the bulk of cognition. But even the greats could not agree. Jean Piaget, for instance, brushed Wundt's view aside. Inhelder and Piaget (1958) proposed that the mental structures required to process experience develop in a stage-like progression from infancy to adolescence. Once children reach the highest stage, they possess "Euclid's understanding of geometry, Newton's ... understanding of space, time, and causality, and Kant's understanding of logic" (Flanagan, 1991, p. 145). For these developmental psychologists, logic was a key descriptive and normative foundation of the mind's highest stage of reasoning. Moreover, cognitive psychologists and scientists, from Bruner (Bruner, Goodnow, & Austin, 1956) to Fodor (1975) to Evans (1982), took theory testing based on deductive logic, thus following Popper (1959/2005), as the key to human learning. When Wason (1969) examined adults' reasoning and found discrepancies from the rules of logical deduction, he and other contemporary cognitive psychologists did not challenge the normativity of logic but inferred that something in his carefully constructed selection task "predisposes people to regress temporarily to less sophisticated modes of cognitive functioning" (p. 478).

Yet Wundt's (1912/1973) rejection of logic as the foundation of cognition experienced a renaissance in psychology, although on the basis of very different arguments. Specifically, the normativity of logic came under attack in two ways in the late 1980s and 1990s. According to Cosmides (1989), natural selection did not evolve general-purpose cognitive algorithms but rather cognitive algorithms that succeed in solving recurrent adaptive problems, such as the threat of being cheated in a social exchange. From this perspective, reasoning obeys a Darwinian and not a formal deductive logic. The second challenge arose in terms of a probabilistic approach being taken to purportedly logical reasoning tasks (Oaksford & Chater, 1994). According to this view, the conclusion that humans reason irrationally results from comparing "apparently irrational behavior ... with an inappropriate logical standard" (Oaksford & Chater, 2001, p. 349). Specifically, people's reasoning is better understood in the Wason selection task in terms of a process of inductive hypothesis testing (and a Bayesian model of optimal data selection) than in terms of an "outmoded falsificationist philosophy of science" (Oaksford & Chater, 1994, p. 608). Consequently, probability theory rather than logic should be the normative benchmark. It is, of course, not without irony here that human reasoning has also been famously observed to deviate from the norms of probability theory (Barbey & Sloman, 2007; Kahneman, 2011; Kahneman & Tversky, 1972). Yet, like in research on reasoning, both the evidence for people's proneness to errors in statistical reasoning (Peterson & Beach, 1967) and the appropriateness of the invoked probabilistic norms for human rationality (Gigerenzer, 1996) have been hotly debated among psychologists as well.

There has thus been a history of opposing views on whether classical logic should serve as a universal benchmark for human rationality. Similar arguments have been raised with regard to probability, coherence, and other benchmarks of rationality (see Arkes, Gigerenzer, & Hertwig, 2016; Hertwig & Volz, 2013). We believe that now is a good moment for psychologists and philosophers to join forces and work toward a new foundation for the definition of rational cognition. This article represents a first step in this direction, with one author being a philosopher and one a psychologist. We begin by briefly outlining two recent developments in psychology—*ecological rationality* and *descriptivism*—that contribute to the ongoing debate about appropriate frameworks of rational cognition.

## 2. Ecological rationality and descriptivism

One development is the concept of "ecological rationality" (Arkes et al., 2016; Gigerenzer, Todd, & the ABC Research Group, 1999; Hertwig, Hoffrage, & the ABC Research Group, 2013; Hertwig, Pleskac, Pachur, & the Center for Adaptive Rationality, in press; A. Kozyreva & R. Hertwig, unpublished data; Todd, Gigerenzer, & the ABC Research Group, 2012). This view endorses the premise that rationality is evaluated against some benchmark but argues that contrary to a frequent assumption in psychology, there are no *universal* benchmarks. What are treated as universal benchmarks—for instance, consistency, coherence-based rules, modus ponens, or Bayes's rule in probability theory—do not suffice to evaluate behavior as rational. Instead, rationality should be measured in terms of the organism's success—accurate predictions or competitive decisions—in the world. Ecological rationality thus aims to shift the researcher's methodological strategy from the a priori imposition of content-free norms to studying the organism's goals and achievements within the context of specific environmental structures as well as the mind's undeniable cognitive constraints. Researchers would thus ask: Under what environmental structure is a given cognitive strategy (e.g., heuristic, rule, routine) for the task at hand more accurate than competing strategies that need more information and computation, and under what structure is it not? A strategy is ecologically rational to the degree that it is adapted, in the context of the task, to the informational and statistical structure of an environment. This also means that any strategy is no longer good or bad, rational or irrational per se, but rather it is or is not adapted to the specific task and environment. In addition, it means that a strategy is commonly being tested against some other strategies that may or may not be even better adapted to the specific task and environment (e.g., Gigerenzer & Brighton, 2009; Spiliopoulos & Hertwig, in press).

Although Elqayam and Evans (2011) classified the concept of ecological rationality among nonnormativist positions, they criticized it for being in danger of committing the dubious inference from "is" to "ought" (see also Elqayam & Over, 2016, p. 46). The position they advocate, *descriptivisim* (Elqayam & Evans, 2011; Elqayam & Over, 2016), is meant to escape the "is–ought" inference trap. The escape is realized by completely eschewing normative concerns. Elqayam and Over proposed that

> the psychology of reasoning and decision making would be better off letting go of normative concerns altogether. Instead of measuring rationality by normative standards, the descriptivist position is that rationality should be *measured by the achievement of personal goals.* (Elqayam & Over, 2016, p. 7, emphasis added)

To this end, Evans and Over (1996) proposed a distinction between rationality$_1$, measured in terms of achieving one's goals, and rationality$_2$, measured against a priori normative standards, such as classical logic or probability theory. Rationality$_1$ is postulated to be personal and contextual, resulting in *instrumental rationality*, meaning that an individual behaves in such a way as to achieve his or her personal goals (see also Elqayam, 2012).

In our view, the opposition between descriptivism (rationality$_1$) and normativism (rationality$_2$) that Elqayam and Evans (2011) invoked is misleading because the character of "instrumental rationality" is ambiguous (for other critical objections see Hertwig, Ortmann, & Gigerenzer, 1997). Ordinarily, the assertion that an action is instrumentally rational means that it is rational because it is the appropriate *means* for some *end* that, in turn, is assumed to be of value. Thus understood, instrumental rationality does involve a normative dimension insofar as it shifts the normative weight from the end to the means (i.e., to the action; Schurz, 1997, sect. 6.1). There is a second, purely descriptive understanding of instrumental rationality according to which the proposition that an action is instrumentally rational for a given end simply means that the action is appropriate for reaching this end, even if this end is *bizarre* from a commonsense or intuitive viewpoint. For example, it would sound odd to describe "heavy smoking" as instrumentally rational in regard to the goal of increasing the chances of developing lung cancer or frequent casual sex as instrumentally rational in regard to the goal of contracting a sexually transmitted disease. Yet such descriptive statements would be perfectly fine in this second understanding.

Notwithstanding this criticism, the notions of ecological and instrumental rationality and descriptivism have one thing in common: They object to the reduction of rationality to allegedly universal normative systems, which are, in turn, founded on a priori intuitions that are inaccessible to further justification in terms of their cognitive functionality or success in the world. Next, we turn to the difficulties such intuitions face. To this end, let us dip our toes in some philosophical waters.

## 3. The problems of justifying rational cognition from a priori norms or intuitions

To appreciate the full force of the foundational issues in question, it helps to briefly consider normative ethics and more specifically the classical distinction between *deontological* and *consequentialist* justifications of ethical norms (see Broad, 1967; Frankena, 1963). In deontological frameworks, the justification of norms is rooted in general a priori intuitions about values and duty principles that are assumed to be "good in themselves" (e.g., Kant, 1785/2011). These principles are obligatory, irrespective of the consequences that might follow from our actions. In consequentialist frameworks, in contrast, how correct our moral conduct is will be determined *solely* by a cost–benefit analysis of the action's consequences. One example of such a framework is (act) utilitarianism, according to which an action is morally justified if the action's total good consequences outweigh its total bad consequences (e.g., Mill, 1863/1998). Let us employ the distinction between deontological and consequentialist justification in the context of rational cognition (see Goldman, 1986, p. 97). As in deontological theories of ethics, in *apriorist* accounts of rational cognition (note that the term *deontological* is reserved for the domain of ethics), norms are justified by reference to a priori intuitions. Such foundational intuitions could be, for instance, necessity, consistency, or coherence. Generally speaking, a norm or an intuition concerning the rationality of a given cognitive strategy can be described as a priori if either it is considered evident without further justification, or its justification is based on other intuitions that are *independent* of the consequences of this strategy in a given environment. In contrast, *consequentialist* accounts of rational cognition justify their benchmarks in terms of what one could call "cognitive success." This means that these benchmarks acquire "normative legitimacy" through the success of their consequences and not through agreement with some norm such as coherence that is imposed a priori (e.g., transitivity, property alpha, procedural invariance; see table 1 in Arkes et al., 2016).

### *3.1. Equilibrium justifications and the problem of circularity*

In our view, it is highly problematic to enlist a priori intuitions as the foundation for justification of rational norms. Let us explain our concern. After five centuries of failed attempts in the history of rationalist philosophy, including Kant's (1781/1998), to justify principles a priori, there is wide consensus in contemporary epistemology: It is impossible to justify cognitive principles from nothing, which was Kant's understanding of "a priori." Thus, contemporary philosophers in the rationalist tradition have put the *coherence of intuitions* at the basis of rationalist justifications that are considered a priori in the sense explained.[1] The method of justifying a priori intuitions by the coherence with other intuitions has been called, perhaps somewhat euphemistically and sidestepping the term *intuition*, the "method of reflective equilibrium" (Cohen, 1981; Goodman, 1955/1983; Rawls, 1971):

> The key idea underlying this view of justification is that we "test" various parts of our system of beliefs against the other beliefs we hold, looking for ways in which some of these beliefs support others, seeking coherence among the widest set of beliefs, and revising and refining them at all levels when challenges to some arise from others. For example, a moral principle or moral judgment about a particular case (or, alternatively, a rule of inductive or deductive inference or a particular inference) would be justified if it cohered with the rest of our beliefs about right action (or correct inferences) on due reflection and after appropriate revisions throughout our system of beliefs. By extension of this account, a person who holds a principle or judgment in reflective equilibrium with other relevant beliefs can be said to be justified in believing that principle or judgment. (Daniels, 2018, sect. 1)

From a consequentialist viewpoint, however, there is a vigorous objection to such "equilibrium justifications." They are *circular*. In reply to this objection, several philosophers have argued that even circular justifications may have epistemic value (e.g., Goldman, 1999, p. 85; Psillos, 1999, p. 82). However, there are strong counterarguments showing that such hopes are in vain. Before turning to one, let us clarify that we do not deny that certain justification structures that have been called circular in the literature can have epistemic value (see Hahn, 2011); yet these are of a different sort from the circles involved in equilibrium justifications.[2]

### 3.2. *Circular justifications and the problem of contradictory intuitions*

One key counterargument to the view that circular justifications have epistemic value demonstrates that *contradictory* rules can be pseudojustified by the same circular argument structure. For example, the circular inductive justification of induction goes as follows: Inductions were successful in the past, whence, by induction, they will be successful in the future. If one accepts this justification, then—to avoid inconsistency—one must equally accept a *counterinductive* justification of counterinduction[3] that runs as follows: Counterinductions were not successful in the past, whence by counterinduction they will be successful in the future (see Douven, 2011, sect. 3; Salmon, 1957; Schurz, 2018). Eventually, circular justification may also be given for fundamentalist doctrines, such as the "rule of blind trust in God's voice," which a person may hold in reflective equilibrium with the intuition that "God's voice in me tells me that I should blindly trust his voice."

The fact that equilibrium justifications can easily support contradictory intuitions demonstrates that circular justifications are highly problematic. Because of their circularity, these contradictory intuitions cannot be meaningfully correlated with the world but are rather inescapably *subjective* in nature. A striking example of an intuition-based account of rationality in psychology and cognitive science is Cohen's (1981) article "Can Human Irrationality Be Experimentally Demonstrated?" According to *Google Scholar*, this article has been cited a total of 1,414 times (June 23, 2018). The philosopher Cohen vehemently argued against the bleak implications for human rationality implied especially by the research in psychology on probabilistic reasoning (Kahneman & Tversky's heuristics-and-biases program; Kahneman, 2011) and deductive reasoning. For Cohen, rules of logical and probabilistic reasoning such as modus ponens, modus tollens, and Bayes's theorem are based on intuitions about correct reasoning. He put it as follows: "The presence of fallacies in reasoning is evaluated by referring to normative criteria which ultimately derive their own credentials from a systematization of the intuitions that agree with them" (1981, p. 317). From this follows, Cohen argued, that if people's reasoning deviates from such rules, then this merely means that they have different intuitions about correct reasoning than logicians or probability theorists do, and therefore experimenters "risk imputing fallacies where none exist" (1981, p. 330).

The subjective nature of intuition-based justifications raises the problem of how to arbitrate between competing normative systems. Some have diagnosed this arbitration problem as essentially unsolvable because cognitive norms, goes the argument, will necessarily be based on intuitions, without external standards of cognitive success (Elqayam & Evans, 2011). Consequently, an intuition-based justification of rationality is doomed to result in a strong form of *cognitive relativism* ("anything goes")—a position whose consequences the philosopher Stich (1990) worked out.

### 3.3. Why a consequentialist account of rational cognition is indispensable

What follows from this discussion? First, we do not deny that intuitions are needed in some areas, for example, in ethics where one inevitably must define what counts as intrinsically valuable. However, the appropriateness of cognitive systems should be evaluated not by intuitions but, so we argue, by demonstrations that these systems have successful consequences in the real world. Cognitive success is thus a concept that brings a consequentialist perspective to the justification of norms for rational cognition. Remember consequentialism (as used in ethics) means that the moral rightness of an act depends only on the consequences of that act. By analogy, an act of a cognitive system is rational insofar as its consequences bear success. For instance, the validity of the rule of modus ponens is established *not* "by intuition" but by the semantic proof of its strictly truth-preserving nature: If "$p$" and "$p$ implies $q$" are true, then "$q$" will be true as well, no matter the environment you are in. This justification of modus ponens is consequentialist in nature. Cohen (1981, p. 319) objected that the "if–then" of classical logic deviates from the if–then in natural language. Therefore, according to Cohen (1981, p. 319), *intuitions* need to be invoked to determine the "right" meaning of the conditional. To this argument, the consequentialist reply is that to assume that there is an "objectively right" meaning is a "rationalistic illusion"—there are only more or less cognitively successful meanings and these can change across contexts (see also Hertwig, Benz, & Krauss, 2008; Hertwig & Gigerenzer, 1999). It is well known that the if–then of natural language has a number of different semantic interpretations (cf. Bennett, 2003). The question of which is cognitively most appropriate should be answered not by reference to intuition but by replacing the ambiguous if–then of natural language with semantically well-defined conditionals (e.g., strict, uncertain, indicative, counterfactual) and investigating their cognitive properties. In later sections we investigate the cognitive success of different systems of strict and uncertain conditional reasoning, with surprising results. Investigations of this sort are impossible as long as these systems are merely evaluated and justified on the basis of intuition, in particular since a growing dissent of intuitions has emerged in the area of conditional reasoning (see Pelletier & Elio, 1997).

In sum, taking intuitions as sacrosanct would hinder empirical research and rational criticism. We suggest that the better justification of norms of rational cognition is *consequentialist* in nature. Within a consequentialist account, the severe problems of arbitration, cognitive relativism, and the indeterminate correspondence of intuitions and the world are either removed or less grave—at least so we claim. The reason lies in the promise that all normative systems of reasoning can be measured on a commensurable metric that we call *cognitive success*. What is it and how can it be measured?

## 4. What is cognitive success?

Next, we propose a consequentialist account of rational cognition. Our account is in line with Quine's naturalized epistemology (1960) but goes beyond it in its explication and applications of the notion of cognitive success, as well as in its new understanding of the interplay between its descriptive and normative components. What distinguishes the present proposal from all naive sorts of pragmatism is that cognitive systems are evaluated in terms of *cognitive* rather than *practical* success indices (such as moneymaking). What is measured by cognitive success is the "cognitive part" of rationality. Cognitive rationality is a precondition for practical rationality, but unlike practical rationality, it abstracts from the question of what ends are normatively right or intrinsically good. In contrast, practical rationality, in the philosophical understanding of this concept, attempts to answer this question. For example, knowing the optimal temperature for roasting meat is "cognitively rational," but the assessment of the practical rationality of roasting meat depends on one's ethical attitude toward a vegetarian versus nonvegetarian diet.

A consequentialist approach to the definition of rational cognition faces two main challenges. First, how can the value of cognitive success be justified without again presupposing normative intuitions, thus inheriting all the problems outlined above? According to philosophical arguments harking back to Hume (1739/40) and Moore (1903), it is impossible to derive norms solely from the "is," that is, from empirical facts (Schurz, 1997). Consequently, every instrumental justification of particular norms must assume, besides factual information, more general norms. For example, inferring that calisthenics is good from the fact that it improves fitness assumes that fitness is a general norm. Does this argument then not thwart any attempt to ground the notion of cognitive success in anything but, again, normative intuitions?

Although this objection—every instrumental justification of particular norms must assume more general norms—is logically correct and has useful applications in ethics (Schurz, 2014), we argue that it does not apply to psychology and cognitive science for the following reason: *Cognitive success is instrumental for all—or at least most*[4]*—purposes*. Every real-world decision problem involves, as a part of it, a ubiquitous cognitive task, namely, predicting which of the available actions will have the maximum expected payoff, in light of a given reward function.[5] Greater success in this cognitive task will, by and large, lead to greater success in one's actions, independently of the goals pursued (Schurz, 2014). Is the premise that cognitive success is instrumental for almost all purposes really sufficient for the normative justification of cognitive success? Logically speaking, *no*, because this premise is descriptive and (as explained above) no "ought" can follow from an "is" by rules of logic alone. However, the missing normative premise that fills the logical gap is relatively harmless: We assume that it is by-and-large good to help people attain their personal goals. This is indeed a fundamental and widely shared intuition, though not a cognitive but a moral one.

Moreover, the insight that cognitive success is instrumental for almost all practical purposes helps solve the problem of the apparent *relativity* of instrumental rationality to one's assumed purposes, which for many authors constitutes a fundamental obstacle to the objectivity of this notion (e.g., Stich, 1990, p. 131). We suggest that the *purpose-invariant* core of all forms of instrumental rationality is precisely their cognitive rationality (Kornblith, 2002, p. 158). Thus, there are no separate forms of instrumental rationality for cooks, clerks, and pilots, or for right-wing and left-wing politicians. What is common to all these applications of instrumental rationality is their cognitive success. This means that cognitive success is not to be mistaken for moral rightness.

This brings us to the second challenge to a consequentialist approach to defining rational cognition: the meaning of cognitive success. The details will depend on the cognitive task at hand. Yet there must be a core meaning of "cognitive success" that is common to all competing systems of rational reasoning; otherwise, it would be impossible to compare them using the same currency. Above, we argued that every real-world decision problem involves—or can be reformulated in terms of—some kind of *prediction* problem. On the basis of this premise, we suggest the following definition:

> The core meaning of the cognitive success of a system (including algorithms, heuristics, rules) is defined in terms of successful predictions, assuming a comprehensive meaning of prediction that includes, besides the predictions of events or effects, predictions of possible causes (explanatory abductions) and in particular predictions of the utilities of actions (decision problems).

Characterizing a decision problem in terms of a prediction task might seem narrow. Yet much of what people do is predicated on implicit or explicit forecasts about how the future will unfold. Choosing a job, getting married, having children and investing in their education, purchasing an apartment, voting for a party, saving for old age, choosing a medical treatment—all these decisions and many others are reached on the basis of predictions about what the future holds. Moreover, focusing on predictions by no means implies that important cognitive processes are ignored. Since reliable predictions are based on an inductive inference from *sufficiently informed* premises, they engage various nonpredictive subprocesses such as search, memory retrieval, and language processing. Importantly, the major purpose of the predictive reformulation of decision tasks is to *measure* their cognitive success on a commensurable scale. For example, consider the decision problem of buying the "best" car (relative to the buyer's preferences) where the buyer encounters two websites offering two competing decision methods, $M_1$ and $M_2$, to potential car buyers. Then the claim that method $M_1$ is more appropriate for a certain group of car buyers (e.g., males between the ages of 20 and 30) amounts to the testable prediction that the degree of future satisfaction of car buyers in this group, having used method $M_1$, is significantly higher than those who used method $M_2$.

Upon closer inspection, the predictive success of a cognitive system or (more generally) a cognitive method depends on two components that are commonly in competition and whose optimization thus involves a trade-off. In the psychological literature, this trade-off is reflected in the distinction between the *ecological validity* of a prediction method (Brunswik, 1952; Gigerenzer et al., 1999) and its *applicability*.[6] More precisely, a method's cognitive success can be factorized into the *product* of these two components as follows:

$$\text{cognitive success} = \text{ecological validity} \times \text{applicability}$$

where applicability is the percentage of targets for which the method renders a prediction, among all intended targets of prediction, and ecological validity is the sum of scores divided by the number of all predictions rendered, and

$$\text{score (per prediction)} = \max - \text{loss}$$

where max is the maximal score that a perfectly accurate prediction can obtain and loss is a monotonically increasing function of the distance between the predicted and the actually observed value of the event variable. From this it follows that

cognitive success = sum of scores divided by number of all intended targets of prediction.

Ecological validity and applicability of a cognitive method are in competition. One can increase the ecological validity of a method by having it apply only to those few target domains for which the method's predictions are known to be accurate because, for instance, the method has been fitted to this domain. Likewise, one can increase the applicability of a method by applying it also to target domains for which its error rate remains unknown or even known to be high, or by permitting the method to make a random guess in cases where the algorithm does not reach a decision (i.e., in this sense is not applicable). Also note that the definition of the concept of applicability is related to all *intended* target domains but not to all possible target domains. Thus, a method's cognitive success cannot be deemed to be low because it does not apply to domains that were never intended to be part of the class of target domains. Consider, for illustration, the analogy of a hammer—its "success" is not diminished by the fact that the hammer is not suitable to drill holes. We also emphasize that a method's class of intended targets domains is not an invitation to propose arbitrary reference classes but rather is empirically inferred in terms of the method's purposes across all users. Thus, a method's cognitive success cannot be arbitrarily boosted by winnowing down its intended targets to "easy ones."

The score that a method earns for each prediction is its maximally achievable score (max) minus its distance to the observed value (loss). The type of loss function[7] and max are specified by type and context of the given task.[8] Often max is identified with the greatest possible loss; this entails that min, that is, the minimal score, is zero. If loss is identified with the absolute distance function, max is given as the width of the observable value range. For example, if the task consisted in forecasting the next day's mean temperature with values lying in the range between −20°C and +40°C and the loss function is given as the absolute difference between predicted and actual mean temperature, then max is 60°C. If the task is the prediction of probabilities such as that it will rain tomorrow, then, according to a famous result of Brier (1950), the appropriate loss function is not the absolute but the squared distance between the predicted probability and the truth value of the predicted event; thus, max = 1 (true) and min = 0 (false). In the example of people intending to buy a car, a natural loss function might be the absolute difference between mean degree of satisfaction (in an unbiased sample) with the car type recommended by method $M_1$ and that recommended by method $M_2$, with degree of satisfaction measured on a scale ranging, say, from min = 0 to max = 10.

### 4.1. Some possible objections to cognitive success

Let us freely admit that intuitions can play a role in determining the details of the scoring function. However, robust results should be largely invariant to changes of the scoring functions (see the section on uncertain conditionals below). Another objection to the concept of cognitive success is that it downplays the role of explanations relative to predictions. This challenge can serve as a further test case for our account. Salmon (1984) argued that what distinguishes explanations from predictions is that, whereas predictions can be based on noncausal correlations, explanations must spell out the *causes* of the event to be explained. Although we agree, we emphasize that causality can easily be embedded into the concept of cognitive success. What distinguishes a causal from a noncausal correlation between a variable $X$ and another one $Y$ is that the effect of an intervention on $X$ will be transmitted to $Y$ only if $X$ is a cause of $Y$ (this is a consequence of the causal Markov condition; see Pearl, 2009). Thus, the cognitive success of causal information resides in its capacity to *predict* the consequences of (human) actions.

Another account identifies good explanations with argument patterns that *unify* many empirical phenomena (Kitcher, 1981). However, it can be shown that empirical unification correlates with empirical confirmation and this, in turn, correlates with predictive success (Schurz & Lambert, 1994). The only notions of explanation that are not and *should* not be covered by our account are those that make the quality of an explanation dependent on its coherence with “intuitions of understanding” and that are inexplicable in terms of causal or unificatory concerns.

The two core components of our notion of cognitive success, ecological validity and applicability, are related to a number of further important evaluative dimensions:

- A method with high ecological validity has a high *truth rate*[9] in those situations where it is applicable; thus, high ecological validity is connected with low risk of error.
- A method with high ecological validity may nevertheless have low cognitive success if it can rarely be recruited due to low applicability.
- A method with high applicability renders predictions possible across many predictive contexts. High applicability therefore suggests that the method has a high information output.
- On the other hand, a method’s applicability is inversely related to its cognitive costs, measured in terms of the information input needed and the effort required to process it. The higher the cognitive cost of a method, the more often it will be inapplicable because it exceeds the upper bound of agents’ cognitive resources (see also Payne, Bettman, & Johnson, 1993).

The threefold tensions between risk of error, information output, and cognitive costs create a fitness landscape[10] that can explain many facets of the pros and cons of competing systems of rational reasoning. How these cognitive fitness factors interact in concrete cognitive tasks will be discussed next. In particular, the tension between these factors explains why cognitive science needs not a monism but a *pluralism* of cognitive methods, and why the evaluation of those methods' advantages and weaknesses should rely not on intuition but on careful comparison of their respective success. Next, we illustrate this point by applying the notion of cognitive success in the domains of classical material conditionals, uncertain conditionals, Bayesian probabilities, and prediction and choice.

### *4.2. Cognitive success and deductive reasoning*

Let us return to classical logic, our introductory example of what many psychologists considered a universal norm of rational cognition in the 20th century. Deductive inferences are, by definition, completely valid—that is, they have maximum ecological validity (1.0): In *all* situations in which all premises are true, the derived conclusion will invariably be true. Yet this ideal validity of deductive inferences stands in stark contrast to their very low applicability, as emphasized by Wundt (1912/1973; see above). That is, the prevalence of deductive inferences with *nontrivial* conclusions is low. As an example, consider inferences of propositional logic involving the classical (material) conditional "If *P*, then *Q*" (semantically equivalent to "not-*P* or *Q*"). It can be shown that this inference can have a nontrivial conclusion insofar as it is possible to confirm each premise by observations that do not already contain the conclusion. This will be the case if the following condition is satisfied: The verification of the conditional premise "If *P*, then *Q*" is based not on the observation of "not-*P*" or of "*Q*," but rather on an inductively supported belief that expresses (at least implicitly[11]) a *strict* (exceptionless) generality of the form "For all *x* in a given domain: If *P*(*x*), then *Q*(*x*)" (see Schurz, 2014, sect. 5.1). Exceptionless regularities (i.e., conditional probabilities of 1.0) are known to be rare in empirical (nonmathematical) domains. What does this mean? It simply means that inferences of propositional logic with nontrivial conclusions are rare. Therefore, their overall cognitive success will be low in these domains, notwithstanding their maximum validity. Only if one could demonstrate that in a specific environment the applicability of deductive reasoning is high could one argue in favor of this system's high cognitive success in this environment. One such environment may be cheater detection, where people can be, under specific circumstances, remarkably successful when measured in terms of modus tollens reasoning (Cosmides & Tooby, 1992). Another domain may be consistency checks in legal reasoning (Arkes et al., 2016).

### *4.3. Cognitive success and reasoning with uncertain conditionals*

Uncertain conditionals are conditionals of the form "If *A*, then normally *B*." They are epistemically acceptable if the associated conditional probability pr($B|A$) is "sufficiently" high, that is, higher than a contextually determined threshold $\alpha > .5$. Systems of probability logic infer further conditionals from sets of uncertain conditionals. There are four well-known systems of reasoning with uncertain conditionals: O, P, Z, and QC. System O (Hawthorne & Makinson, 2007) is the only system that preserves epistemic acceptability from premises to conclusion for *any* chosen acceptability threshold. System P is the famous system of probability logic developed by Adams (1975). It guarantees to preserve epistemic acceptability only if the sum of the premises' conditional uncertainties is smaller than 1.0 minus the acceptability threshold (where uncertainty is defined as 1.0 minus probability; Oaksford & Chater, 2007, p. 111). System Z goes back to Pearl (1990) and makes additional default assumptions that, roughly speaking, maximize the entropy of the distribution under the high-probability constraints dictated by the premise conditionals (Hill & Paris, 2003). System QC (for "quasi-classical reasoning") reasons with uncertain conditionals as if they were exceptionless conditionals of classical logic.

For illustration, assume a small world with only four predicates: "being a bird" (B), "being able to fly" (F), "having wings" (W), and "being male" (M). The known premises are the two uncertain conditionals (a) $B \Rightarrow F$ (birds can fly) and (b) $B \Rightarrow W$ (birds have wings), with associated probabilities $pr(F|B) = pr(W|B) = .95$. System O draws only trivial inferences such as $B\&(M \vee \neg M) \Rightarrow F$ from Premise a, meaning birds that are either male or not male can fly, with an associated probability of .95. In addition to the previous inference, system P draws the inference $B\&W \Rightarrow F$ from Premises a + b, meaning, birds having wings can fly, with an associated probability of .9. System P does so by applying the law of "cautious monotonicity" and the uncertainty sum rule. In addition to the previous inferences, System Z draws the inferences $B\&M \Rightarrow F$ and $B\&\neg M \Rightarrow F$ from Premise a, meaning male birds as well as nonmale birds can fly, with an associated probability of .95. It does so by making the default assumption that the predicates "male" and "being able to fly" are statistically independent (likewise in application to Premise b). Finally, in addition to all previous inferences, System QC draws the "risky" inference of contraposition $\neg F \Rightarrow \neg B$, meaning nonflying objects are not birds. This follows from Premise a with an associated probability of .95 (similarly in application to Premise b).

These four systems differ significantly in their predictive power. They become increasingly powerful and, at the same time, more risky and error prone. That is, the applicability (number of derived conclusions) and error probability (number of mistakes made) increase from O to P to Z to QC. From a consequentialist viewpoint, the question is not which of these systems is the right or true one, but which is superior with regard to cognitive success. Schurz and Thorn (2012) performed a cognitive-success analysis of the Systems O, P, Z, and QC. In their computer simulation an environment with four binary variables *a*, *b*, *c*, *d* and a randomly generated probability distribution was repeatedly simulated. The possible cases (predictive targets) consisted of all 464 conditionals with conjunctions of one, two, or three unnegated or negated variables in their antecedent or consequent. The task on which the four systems were compared was the derivation of conditionals from four randomly selected base conditionals with conditional probabilities $\geq .7$, together with a prediction of their associated conditional probabilities.[12] Thus, there were at most 460 conditional probabilities to be predicted. Four different scoring rules for cognitive success were compared. Table 1 presents the results for the ACG (advantage compared to guessing) score, defined as the absolute difference between the predicted and the actual conditional probability for each of the derived conditionals. Although the ordering of the four systems according to their ecological validity is Q > P > Z > QC, their applicability ordering is precisely the inverse, QC > Z > P > Q. The resulting cognitive success ordering is Z > QC > P > O.

In light of these results, Schurz and Thorn (2012) concluded that System Z achieves the optimal balance in the trade-off between deriving true and informative conclusions and avoiding false or uninformative ones.[13] Schurz and Thorn (2012) and Thorn and Schurz (2014) investigated three additional scoring rules: PIR (price is right), sPIR (subtle price is right), and EU (expected utility). The qualitative orderings of the ecological validity, applicability, and cognitive success of the four systems were the same across all four success measures, demonstrating the robustness of the results.[14]

Table 1
Cognitive success analysis of four systems of reasoning with uncertain conditionals

| System | Applicability (% of 460 intended predictions) | Sum of Scores (ACG score)[a] | Ecological Validity (range [0, 1]) | Cognitive Success (range [0, 1]) |
|---|---|---|---|---|
| O | 1.0 | 4.6 | 0.92 | 0.009 |
| P | 1.4 | 5.2 | 0.82 | 0.011 |
| Z | 10.5 | 22.5 | 0.47 | 0.049 |
| QC | 22.9 | 8.5 | 0.08 | 0.018 |

*Note*. [a]For normalization purposes, the ACG scores in table 2 of Schurz and Thorn (2012) were multiplied by 3.

### *4.4. Success and Bayesian probabilities*

Bayesian probabilities are internally coherent degrees of subjective beliefs. Following arguments by Ramsey (1926/1990) and De Finetti (1937/1964), coherence is usually justified as follows: If one interprets degrees of beliefs as fair betting quotients, one is guaranteed to never accept a system of bets that exacts a logically guaranteed loss, that is, a "Dutch book."[15] The Dutch book argument is thus indeed a consequentialist justification as it ties the consequences of a person's subjective probabilities back to monetary outcomes. However, what is thus justified is merely the coherence of probabilities, and this means only that they need to satisfy the basic (Kolmogorovian) probability axioms.[16] This is indeed a necessary constraint on rational degrees of beliefs, but by itself it is not sufficient for rational degrees of belief to yield cognitive success. The condition of a coherent fair betting quotient depends solely on the gambler's subjective beliefs and preferences. It does *not* involve any adaptation to the environment, that is, to the true frequency or statistical probability (frequency limit) of the events betted on. Consider, for example, a subjectivist who repeatedly offers betting odds of 1:1 that she will roll a 6 with an unbiased die. She considers this bet to be fair and is equally willing to accept the opposite bet that she will not roll a 6. She is coherent and will remain coherent even after she has lost her entire fortune. She will be puzzled that while everybody readily accepted her first bet, nobody accepted the opposite bet, even though both are equally fair in her view. Thus, if she ignores the frequentistic chances of the events betted on, she will be unable to explain why she lost everything and others won.

As this, admittedly engineered, example illustrates, the problem of subjective degrees of belief is not their low applicability (an individual's beliefs could discriminate between many states of the world) but their potentially low ecological validity. The Bayesian coherence requirement is too weak to exclude cognitively *unsuccessful* behavior if one's degrees of belief are not connected with objective truth-chances (i.e., statistical probabilities; Knight, 1921). There are pertinent methods in Bayesian statistics of establishing this connection (less well-known than the Dutch book arguments), such as Lewis's "principal" principle (1980/1986) or De Finetti's (1937/1964) equivalent "exchangeability" principle. These principles demand that a person's rational degree of belief (Pr) in an event (E) should have the value $r$, given that all that the person knows is that the statistical probability (pr) of the corresponding event type E is $r$, more formally, $\Pr(E \mid pr(E) = r) = r$ (for arbitrary $r \in [0, 1]$). One can prove that the satisfaction of this principal principle is equivalent to the assumption of Bayesian statistics that degrees of belief can be represented as weighted averages of statistical probabilities. Subjective probabilities that satisfy this condition are known to converge toward the true statistical frequencies when the evidence increases infinitely, independently of the assumed prior distributions (Gillies, 2000, p. 71ff; Howson & Urbach, 1996, chapter 14; Schurz, 2013, pp. 165, 236f). It is only if this connection between subjective and objective probabilities is established that Bayesian reasoning can be cognitively successful and decisions based on maximization of subjectively expected utility can maximize one's average utility.

### *4.5. Cognitive success in prediction and choice*

Perhaps more than in any other research area in psychology the tension between apriorist and consequentialist accounts of rational cognition unfolded in the debate about the meaning of bounded rationality in general and the role of heuristics in particular. The heuristics-and-biases research program (Kahneman, 2011), possibly the most influential research program in psychology of the last five decades, has consistently invoked the rules of probability theory and statistics as a priori norms for human rationality. Deviations from these norms in people's reasoning were taken as manifestation of irrationality. In Kahneman's (2003) portrayal of the program's research, it "attempted to obtain a map of bounded rationality, by exploring the systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models" (p. 1449). Many of the systematic biases were attributed to the operation of heuristics (e.g., availability, representativeness, and anchoring-and-adjustment) that although "quite useful," sometimes "lead to severe and systematic errors" (Tversky & Kahneman, 1974, p. 1124). On this view, a heuristic's rationality is evaluated exclusively on the basis of its conformity to the norms and not in terms of its potential cognitive success.

This changed with the arrival of the ecological rationality research program, which has redefined the normative study of heuristics; by extension, it interprets bounded rationality in terms of the match between a heuristic and an environment, the two blades in Simon's (1990, p. 7) scissors metaphor. On this view, this match determines the performance and thus the cognitive success of a heuristic. In order to measure cognitive success, researchers of heuristics' ecological rationality have conducted a wide range of tournaments between simple heuristics and complex strategies commonly considered to be normative. These computer simulations encompass, for instance, the analysis of heuristic inferences about real-world quantities (e.g., which of two cities has a larger population size; Gigerenzer & Brighton, 2009; Gigerenzer & Goldstein, 1996; Katsikopoulos, Schooler, & Hertwig, 2010) and more recently the analysis of choices between uncertain lottery options (Hertwig, Woike, Pachur, & Brandstätter, in press) and of choices in strategic games (Spiliopoulos & Hertwig, in press). For illustration, consider the tournament involving choice strategies choosing between uncertainty lottery options (Hertwig et al., in press). The simulations implemented 20 choice environments (defined by different payoff and probability distributions) and randomly generated 6,000 choice problems per environment. The innovation in this simulation was that all strategies (with the exception of the omniscient expected value model) learned about the properties of each problem by sequentially taking one draw at a time from each of the options per problem. The strategies then chose what they inferred to be the best option after each sample (learning stopped after 50 rounds).

Table 2 presents the cognitive success of each of the six choice strategies. The normative benchmark for human beings is either the omniscient expected value theory, or, more realistically, the sampling-based expected value theory. In light of the cognitive success measure, Hertwig et al. (in press) concluded that under uncertainty (when all strategies have incomplete knowledge and need to sample the environment), some simple choice heuristics nearly approximate the performance of the sampling-based expected value theory—even though they may not take entire swaths of information into account. The well-performing equiprobable heuristic, for instance, ignores all probabilities and merely calculates the mean of all outcomes within each option, then chooses the option with the highest mean. Indeed, the research on ecological rationality has repeatedly demonstrated that simple heuristics, which curtail search for information and reach decisions without complex calculations, can lead to surprisingly good inferences and predictions relative to complex algorithms based on the principles of logic, probability theory, and maximization.

The strategies in Table 2 selected an option randomly in cases where their policy and the information available did not render a choice, that is, were not applicable. For this reason their applicability is always 100% and their ecological validity and cognitive success are identical. In other tournaments measuring cognitive success some competing methods have low applicability, whereas others are always applicable. This is particularly the case in tournaments including *meta-inductive* selection strategies. The account of meta-induction (Schurz, in press; Schurz & Thorn, 2016) is in an important sense complementary to the research program of ecological rationality: Meta-induction is a general meta-cognitive strategy designed to choose, in each situation in which it is applicable, a locally optimal method from a given toolbox of candidate methods. Two important meta-

Table 2
Cognitive success analysis of choice strategies in choice environment requiring learning of the properties of the choice options

| Strategy[a] | Applicability (in %)[b] | Cognitive Success (range [0%, 100%])[c] | | |
|---|---|---|---|---|
| | | $N$ = 5 | $N$ = 20 | $N$ = 50 |
| Equiprobable | 100 | 93.1 | 94.6 | 93.7 |
| Probable | 100 | 86.4 | 92.3 | 93.5 |
| Lexicographic | 100 | 86.4 | 87.9 | 88.0 |
| Least-likely | 100 | 54.2 | 61.5 | 64.3 |
| Sampling-based expected value theory[d] | 100 | 94.0 | 98.3 | 99.3 |
| Omniscient expected value theory | 100 | 100 | 100 | 100 |

*Notes.* [a]All strategies are described in detail in Hertwig et al. (in press).
[b]In this analysis all strategies were always applicable because they could either select the options or choose randomly.
[c]Average performance across all 20 choice environments and for $N$ = 5, 20, and 50 samples taken per option (two options with two, four, and eight outcomes) from the environment; the cognitive success metric is normalized such that 100% means that a strategy always selected the option with the higher expected value (as did the omniscient expected value model) and 0% means that a strategy always selected the option with the lowest expected value.
[d]The sampling-based expected value theory can also be implemented in terms of a simple heuristic (i.e., natural-mean heuristic; see Hertwig et al., in press).

inductive strategies are *take-the-best*[17] and *success-weighting*. Take-the-best applies in each round of the tournament. It selects the prediction method that is applicable (i.e., renders a prediction) and that has the best success record in the past. Success-weighting predicts a weighted average of the predictions of those methods that rendered a prediction in the given round of the tournament, with weights reflecting the methods' past successes. Table 3 presents the results of applying take-the-best and success-weighting to the results of the Monash University footy tipping competition (MUFTC[18]). The predictive target was forecasting the 3-valued results (1, 0, or tie) of matches of the Australian Football League. The tournament included the predictions of 1,071 human participants (Table 3 reports the five human forecasters with the highest success rates, Forecasters 1–5) as well as the predictions of the different meta-induction strategies including take-the-best and success-weighting.

The five best human forecasters displayed high performance only in certain rounds and refrained from making predictions in other rounds. The meta-inductive strategies utilized the predictions of the best human forecaster in each round, with the result that their applicability was 100% and their cognitive success surpassed that of the best human forecasters (with a slight advantage of success-weighting over the simpler take-the-best strategy).

### 4.6. What does cognitive success mean for the is–ought relationship?

Cognitive success' intent to focus on the consequences of rational cognition suggests a new view of the relationship between the normative ("ought") and the descriptive ("is")

Table 3
Cognitive success analysis of the Monash University footy tipping competition (after 1,514 rounds)[a]

| Predictor | Applicability in % | Sum of Scores | Ecological Validity (range [0, 1]) | Cognitive Success (range [0, 1]) |
|---|---|---|---|---|
| Success-weighting | 100 | 877 | 0.579 | 0.579 |
| Take-the-best | 100 | 873 | 0.577 | 0.577 |
| Forecaster 1 | 39 | 839 | 0.640 | 0.554 |
| Forecaster 2 | 27 | 811 | 0.637 | 0.536 |
| Forecaster 3 | 13 | 789 | 0.666 | 0.521 |
| Forecaster 4 | 12 | 789 | 0.676 | 0.521 |
| Forecaster 5 | 13 | 787 | 0.658 | 0.520 |

*Notes.* [a]Target was forecasting the results of 1,514 matches of the Australian Football League over eight seasons from 2005 to 2012. The tournament included the predictions of 1,071 human participants and the predictions of various meta-induction strategies including take-the-best and success-weighting.

dimensions of theories of reasoning. According to the traditional division of labor, it is the task of armchair philosophy to address normative issues, and that of empirical psychology to answer descriptive questions. From the consequentialist perspective of cognitive success, however, empirical results can become normatively relevant, and normative innovations can suggest new empirical questions (see Corner & Hahn, 2013).[19] The relationship between the normative and the descriptive, as conceptualized in different accounts of rationality in reasoning—such as apriorism, descriptivism, or ecological rationality—emerges most clearly from the answers they give to the following question: What should one infer from a *conflict* between *descriptively observed* and *normatively recommended* behavior in the context of a cognitive task?

In general, the answers depend on the theoretical positions adopted. Thus, how scholars respond to the gap between "is" and "ought" is diagnostic with regard to the justification of the rationality norms they endorse. Let us assume that an experiment reveals a divergence between how people reason and how they ought to reason according to some standard rational benchmark such as norms of probability theory, logic, or axioms of rational choice. If a scholar adopts an *intuition-based justification* of rational cognition, the normatively recommended behavior is not defined in terms of its cognitive success but by reference to a priori intuitions. Thus, the cognitive behavior observed will be judged to be irrational. Alternatively, a scholar may endorse a strong relativism of different systems of intuition. In this case, however, no strong rationality inferences can be drawn (Cohen, 1981; Shier, 2000, p. 78).

In *consequentialist* accounts, in contrast, both the empirically observed reasoning and the "reasoning" of the normative systems will be evaluated with regard to cognitive success. For example, if logistic regression were regarded as the normative standard for predicting the value of a criterion based on a set of cues, then the cognitive success of this normative standard, assuming some statistical knowledge base, could be measured against the cognitive success of people's predictive inferences from the same input (see also Gigerenzer & Brighton, 2009). This opens up a new option for responding to conflicts between empirical observations and normative recommendations. The cognitive success of observed reasoning is not necessarily worse than that of the normative system. Indeed, observed reasoning may, in fact, even outperform normative recommendations. As mentioned before, evidence for the latter has been compiled in research on bounded and ecological rationality (Gigerenzer & Brighton, 2009; Gigerenzer, Hertwig, & Pachur, 2011; Hertwig et al., 2013, in press; Todd et al., 2012). If the observed cognitive behavior outperforms the normative system, the consequentialist would need to conclude that the assumed "normative system" is second best and, thus, can no longer be invoked to derive normative recommendations.[20] If, however, observed behavior scored lower on cognitive success than the normative system, the consequentialist's conclusion would depend on a second *theoretical* choice that is open to consequentialists but not to intuition-based accounts, namely, attitudes toward *cognitive adaptationism*. This position assumes that human cognition is near-optimally adapted to its relevant environments. Therefore, a consequentialist proponent of cognitive adaptationism may be inclined to argue that the assumed measure of cognitive success is inappropriate and in need of revision. In contrast, a nonadaptationist consequentialist would conclude, faced with evidence that observed behavior's cognitive success is surpassed by that of the normative system, that human cognition is below par.

Let us explain the relationship between cognitive consequentialism and cognitive adaptationism in more detail. One might think that cognitive consequentialism entails cognitive adaptationism because the former evaluates cognitive systems by their cognitive success and cognitive success entails being well adapted. This reasoning, however, mistakes "is" and "ought." Cognitive consequentialism makes the *normative* claim that cognitive systems *should* be evaluated in terms of their cognitive success. This implies the normative claim that, ceteris paribus, cognitive systems *should* be well adapted to their environment. However, cognitive adaptationism is not a normative requirement but an *empirical* thesis, stating that because humans are the product of evolutionary selection processes, they will be cognitively well adapted. This may or may not be the case—an issue to which we return below. For the present discussion, however, it is important to note that cognitive adaptationism is not entailed by the normative requirement of cognitive consequentialism. Consequently, a cognitive consequentialist can be more or less inclined toward the choice to assume cognitive adaptationism. This choice, in turn, will determine the response to an instance in which actual cognition scores lower on cognitive success than the normative system.

### *4.7. Cognitive consequentialism and the issue of adaptationism*

Let us finally discuss cognitive adaptationism as found in Anderson's (1990, 1991a,b) work in more detail because, prima facie, his rational analysis shares significant resemblances with our account of cognitive consequentialism. Anderson's method consists of five iterative steps (Anderson, 1991a, p. 473):

(1) Specify the goals of the cognitive system.

(2) Develop a model of the environment to which the system is adapted.
(3) Make minimal assumptions about computational limitations, such as memory storage and computation time.
(4) Derive the optimal behavior given in (1)–(3) above.
(5) Finally, test empirically whether the predictions of the optimal behavior derived in (4) are confirmed by human cognitive performance; if not, then the task–environment model developed in (1) + (2) has to be revised.

Anderson has applied rational analysis to domains such as memory analysis, categorization theory, causal inference, and problem solving.[21] Regarding Step 1, in these domains Anderson identifies the goal of the cognitive system as some kind of predictive inference. Thus, there is agreement between Anderson's analysis of cognitive goals and our definition of cognitive success. Also Steps 2 and 3 are consistent with a cognitive success analysis, except that in Step 3 we would argue that "minimal" assumptions should be replaced by *realistic* assumptions about computational limitations (see Simon, 1990). The first crucial difference to our account appears in Step 4. The consequentialist account does not intend to "derive" the optimal method from the description of the task and environment, because apart from very simple cases this is impossible. In the area of prediction methods, the nonexistence of a universally optimal method is the content of Wolpert's (1996) famous *no free lunch* theorem (cf. Schurz, 2017). Simon (1991) demonstrated that what does the real work in Anderson's derivations are specific auxiliary assumptions about the cognitive system and its environment. We mostly agree with Simon's critique of Anderson's optimal adaptation hypothesis: All that rational analysis can do is consider all *available* but not all possible competing methods for a given task and investigate their cognitive success. This is what the consequentialist account suggests.

Cognitive success, unlike optimal adaptation, thus behaves similarly to how Simon portrayed natural selection:

> The theory of natural selection is not an optimizing theory for two reasons. First, it can, at best, produce only local optima, because it works by hill-climbing up the nearest slope. It has no mechanism for jumping from peak to peak. . . . Second, it selects only among the alternatives that are available to it. (1991, p. 29)

This brings us to Anderson's Step 5. This step presupposes the adaptationist thesis that human cognitive behavior is nearly optimally adapted (Anderson, 1991a, table 1, p. 473). As a general claim, such an "evolutionary optimism" is difficult for several reasons. First, evolutionary selection sometimes produces suboptimal and even dysfunctional adaptations (Ridley, 1993, p. 343f). Second, although genetic evolution optimizes the biological reproduction rate, it is less clear how this process relates to cognition. Anderson acknowledged the fact that evolutionary selection does not find a global optimum but merely a local one. However, there is a world of difference between a global and a local maximum: It can be as large as the difference between a sand hill and Mount Everest. All single constraints on cognitive processes dictated through the biological architecture of humans' brains (see Jones & Love, 2011) are concealed in this difference.

In contrast to rational analysis, cognitive consequentialism does not imply cognitive adaptationism, though it is compatible with it. Obviously, the human brain is well adapted in many respects, but not in all. Therefore, we suggest that the view of rational cognition that appears most defensible and productive for contemporary cognitive science is that of a *cognitive consequentialism* that is not bound to strong adaptationist assumptions. In conclusion, the consequentialist account proposes to modify Anderson's Steps 4 and 5 as follows:

(4′) Derive the consequences of the available competing cognitive methods [given the output of (1)–(3)] and test their cognitive success.

(5′) Compare the locally optimal method [i.e., the output of (4)] with the actual human behavior.

- (5′.1) If they agree, recommend the locally optimal method and infer that human cognition is well adapted.
- (5′.2) If they disagree, two cases are possible:
  - (5′.2.1) If human behavior outperforms the locally optimal method, search for better cognitive methods (to thus eventually explain human behavior): Backtrack to (4) and iterate.
  - (5′.2.2) If human performance is worse than the locally optimal method, search for local constraints on the mind's cognitive mechanisms that can explain the disagreement. Backtrack to (3), add these constraints, and iterate. At the same time, recommend the locally optimal method as a *rational improvement* of intuitive human cognition that can be learned by cognitive training.

In other words, at this point cognitive consequentialism potentially has educational implications.

## 5. Conclusion: New questions about rational cognition

The study of human cognition and its rationality seems inseparable from the question of how successful it is. In psychology and economic research, rationality of human cognition has often been equated with coherence, that is, rules for internal consistency, often defined by propositional logic and probability theory (see Arkes et al., 2016). However, for decades, psychologists have disagreed over how well coherence-based normative systems describe human cognition and which coherence-based systems (logic, probability theory, or decision theory) should be granted the status of normative benchmarks for cognition (see our introduction). We have discussed the problems that arise when normative systems are justified by reference to a priori intuitions, as is typically the case. As an alternative and a potential response to some of these problems, we propose a consequentialist account of normative systems. The major tenets of this account are as follows:

> Traditional normativism fails because a priori intuitions are inadequate as justifications of norms of rational cognition.
>
> Traditional descriptivism fails because norms of rational cognition are inevitably needed as benchmarks for successful reasoning.
>
> Norms of rational cognition are better justified from a *consequentialist* perspective, that is, in terms of their cognitive success.
>
> The concept of cognitive success of a cognitive method assumes that all decision tasks can be reformulated as prediction tasks.
>
> Cognitive success is defined as the product of ecological validity and the method's applicability. Determining the available method's cognitive success permits one to compare them on the same scale.
>
> Cognitive consequentialism is related to ecological rationality to the extent that cognitive success depends on the given cognitive task and a specific environment.

We hope and believe that this approach (or a similar consequentialist concept) offers a way to overcome the trite division of labor between the empirical study of the mind and philosophy. This approach raises new and interesting questions. For instance, what does cognitive success imply for research that has drawn strong conclusions about the (ir)rationality of human cognition (e.g., the heuristics-and-biases research program; Kahneman, 2011)? Or, assuming that the success of different cognitive methods depends on specific environments and that no method succeeds in all environments, will there be metamethods that are able to select the best method for the environment in question (see "meta-induction"; Schurz, in press)? Finally, to what extent is cognitive success adaptive in an evolutionary sense, and how does the answer to this question influence the understanding of the relation between *is* (observed cognitive behavior) and *ought* (normatively recommended behavior)? We do not have answers to these and other questions, but we hope that we have convinced the reader that it is timely to ask these questions, thus leaving some skirmishes of the "rationality wars" behind us (Samuels, Stich, & Bishop, 2012) and turning the question of what rational cognition is into a less dogmatic and a more empirical one.

## Acknowledgments

We are grateful to the members of the SPP1516 "New frameworks of rationality" and, in particular, Gabriele Kern-Isberner, Markus Knauff, Wolfgang Spohn, and Michael

Waldmann for many helpful discussions and comments. We are also grateful to Igor Douven, Shira Elqayam, Ulrike Hahn, and Lael Schooler for their important comments and we thank Susannah Goss and Deborah Ain for editing the manuscript.

## Notes

1. We do not exclude the possibility of nonintuitionist a priori accounts. Such accounts propose to revise the conception of "a priori"—as, for example, in epistemic optimality accounts (Schurz, 2018; Schurz & Thorn, 2016).
2. Schurz (in press, chapter 3) showed that circular justification structures are able to have epistemic value only if their circles are merely *partial*, as in "self-dependent justifications" (discussed in Hahn, 2011).
3. Counterinduction infers the opposite of what induction infers from given observations. For example, if Fs have been observed to be Gs in the past, then induction infers that they will be Gs in the future, whereas counterinduction infers that they will be non-Gs in the future.
4. An exception is any situation involving placebo effects: Here the belief in a proposition has a positive effect on its believer even if the proposition (e.g., the medical effect of a pill) is false, whereas the recognition of the true state of affairs may destroy this effect. However, we argue that situations involving placebo effects are rare. Moreover, their cognitive understanding can be instrumentally useful, too, though not in a straightforward way.
5. It is important to note that different objective functions (from reinforcement learning theory) will lead to differences in the reward function.
6. The distinction between ecological validity and applicability has been made for probabilistic cues. Here, we generalize the notion of ecological validity to all sorts of prediction methods. In pair-comparison tasks, the applicability of a cue coincides with its discrimination rate (Martignon & Hoffrage, 1999). Note that the concept of applicability does not refer to the issue of whether an agent desires to employ a method but whether, assuming he or she wants to employ it, it can be employed.
7. Standard loss functions are linear, quadratic (or convex) functions of this distance, assuming that predictions and events are represented by real numbers. For discrete (nonnumeric) predictions, the standard loss function is the zero–one loss, assigning loss 1 to a false and loss 0 to a true prediction. In this case, the sum of zero–one scores (where score = 1 − loss) coincides with the number of correct predictions delivered.
8. By normalization every scale can be transformed into one whose max value is 1.
9. The notion of *truth* is understood as correspondence with the world in the standard (Tarskian) sense. Since the notion of cognitive success is based only on predictions of *empirically measurable* events, as opposed to theoretical posits, the

correspondence-theoretic understanding of truth is unproblematic and not beset with Stich's (1990, p. 130) problem of ambiguous reference.

10. The error–information–cost trio resembles Goldman's (1986) threefold evaluation criteria of the reliability, power, and speed of a cognitive method. An account closely related to our notion of a cognitive fitness landscape is Marewski and Schooler's (2011) notion of cognitive niche.
11. A belief $P(a) \rightarrow Q(a)$ is implicitly general if it is justified by an argument that justifies a corresponding belief $P(a_i) \rightarrow Q(a_i)$ for every individual constant $a_i$.
12. Randomly generated distributions that did not verify at least four such conditionals were aborted.
13. The investigation of these four reasoning systems in application to natural environments is work for the future. Thorn and Schurz (2016) investigated the dependence of the systems' success rate on the *entropy* of the environment, with the result that System Z is optimal in those environments in which entropy is not low.
14. A similar scoring independence has been found in many other applications of cognitive success measures. For example, the optimality of regret-based prediction methods has been demonstrated to hold for all convex scoring measures (Cesa-Bianchi & Lugosi, 2006; Schurz & Thorn, 2016).
15. "Dutch book" is a technical term in probability theory that was introduced by Ramsey and de Finetti (see Gillies, 2000). See also https://en.wikipedia.org/wiki/Dutch_book.
16. Another consequentialist line of justifying the probability axioms points to the fact that they minimize the inaccuracy of degrees of belief if measured by the Brier score (Leitgeb & Pettigrew, 2010a,b).
17. The take-the-best strategy has been intensively investigated in research on ecologically rational heuristics (Gigerenzer & Goldstein, 1996), where it processes probabilistic cues in order to predict which of two objects scores higher on a quantitative criterion; in the meta-induction domain, take-the-best is employed as a meta-strategy for selecting between prediction methods.
18. Recorded at the MUFTC website: http://probabilistic-footy.monash.edu/~footy/.
19. Working out these new interactions between the descriptive and the normative was a major objective of the German Research Foundation's (DFG's) priority program SPP 1516. This manuscript emerged from its authors' collaboration within this program.
20. This is not an is–ought fallacy since cognitive success is assumed to be the consequentialist's core norm.
21. Let us also emphasize that the Bayesian approach to human cognition that Anderson's work set in motion has become a major framework within the cognitive sciences (e.g., Griffiths & Tenenbaum, 2006; Oaksford & Chater, 2001, 2007; Tenenbaum, Kemp, Griffiths, & Goodman, 2011).

## References

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic*. Dordrecht, the Netherlands: Springer.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991a). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471–517. https://doi.org/10.1017/s0140525x00070801

Anderson, J. R. (1991b). The place of cognitive architectures in rational analysis. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 1–24). Hillsdale, NJ: Erlbaum.

Arkes, H. R., Gigerenzer, G., & Hertwig, R. (2016). How bad is incoherence? *Decision*, *3*, 20–39. https://doi.org/10.1037/dec0000043

Barbey, A. K., & Sloman, S. A. (2007). Base rate respect: From ecological rationality to dual processes. *Behavioral Brain Science*, *30*, 241–297. https://doi.org/10.1017/s0140525x07001653

Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford, UK: Oxford University Press.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*, 1–3.

Broad, C. D. (1967). *Five types of ethical theory*. London: UK: Routledge & Kegan Paul.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York, NY: Wiley.

Brunswik, E. (1952). The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1, no. 10, pp. 4–102). Chicago, IL: University of Chicago Press.

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge, UK: Cambridge University Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, *4*, 317–370. https://doi.org/10.1017/s0140525x00009092

Corner, A., & Hahn, U. (2013). Normative theories of argumentation: Are some norms better than others? *Synthese*, *190*, 3579–3610. https://doi.org/10.1007/s11229-012-0211-y

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276. https://doi.org/10.1016/0010-0277(89)90023-1

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York, NY: Oxford University Press.

Daniels, N. (2018). Reflective equilibrium. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Spring 2018)*. Available at https://plato.stanford.edu/entries/reflective-equilibrium/. Accessed November 12, 2018.

De Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg & H. E. Smokler (Eds.), *Studies in subjective probability* (pp. 93–158). New York, NY: Wiley. (Original work published 1937)

Douven, I. (2011). Abduction. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Available at: https://plato.stanford.edu/entries/abduction/. Accessed March 9, 2011.

Elqayam, S. (2012). Grounded rationality: Descriptivism in epistemic context. *Synthese*, *189*, 39–49. https://doi.org/10.1007/s11229-012-0153-4

Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, *34*, 233–248. https://doi.org/10.1017/s0140525x1100001x

Elqayam, St., & Over, D. E. (2016). Editorial: From is to ought: The place of normative models in the study of human thought. *Frontiers in Psychology*, *7*, 628. https://doi.org/10.3389/fpsyg.2016.00628

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London, UK: Routledge & Kegan Paul.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Flanagan, O. J. (1991). *The science of the mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Cambridge, MA: Harvard University Press.

Frankena, W. K. (1963). *Ethics*. Englewood Cliffs, NJ: Prentice-Hall.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*, 592–596. http://psycnet.apa.org/doi/10.1037/0033-295X.103.3.592

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669. https://doi.org/10.1093/acprof:oso/9780199744282.003.0002

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.) (2011). *Heuristics: The foundations of adaptive behavior*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744282.001.0001

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

Gillies, D. (2000). *Philosophical theories of probability*. London, UK: Routledge.

Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. I. (1999). *Knowledge in a social world*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/0198238207.001.0001

Goodman, N. (1983). *Fact, fiction and forecast*. Cambridge, MA: Harvard University Press. (Original work published 1955)

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*, 767–773. https://doi.org/.10.1111/j.1467-9280.2006.01780.x

Hahn, U. (2011). The problem of circularity in evidence, argument, and explanation. *Perspectives on Psychological Science*, *6*, 172–182. https://doi.org/10.1177%2F1745691611400240

Hawthorne, J., & Makinson, D. (2007). The quantitative/qualitative watershed for rules of uncertain inference. *Studia Logica*, *86*, 247–297. https://doi.org/10.1007/s11225-007-9061-x

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of “and”. *Cognition*, *108*, 740–753. https://doi.org/10.1016/j.cognition.2008.06.008

Hertwig, R., & Gigerenzer, G. (1999). The “conjunction fallacy” revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305. https://doi.org/10.1002/(SICI)1099-0771(199912)12:4%3C275::AID-BDM323%3E3.0.CO;2-M

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

Hertwig, R., Ortmann, A., & Gigerenzer, G. (1997). Deductive competence: A desert devoid of content and context. *Cahiers de Psychologie Cognitive*, *16*, 102–107.

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

Hertwig, R., & Todd, P. M. (2000). Biases to the left, fallacies to the right: Stuck in the middle with null hypothesis significance testing. Commentary on Krueger on social-bias. *Psycoloquy*, *11*, 28.

Hertwig, R., & Volz, K. G. (2013). Abnormality, rationality, and sanity. *Trends in Cognitive Sciences*, *17*, 547–549. https://doi.org/10.1016/j.tics.2013.08.011

Hertwig, R., Woike, J. K., Pachur, T., & Brandstätter, E. (in press). The robust beauty of heuristics in choice under uncertainty. In R. Hertwig, T. J. Pleskac, T. Pachur, & The Center for Adaptive Rationality, *Taming uncertainty*. Cambridge, MA: MIT Press.

Hill, L. C., & Paris, J. B. (2003). When maximizing entropy gives the rational closure. *Journal of Logic and Computation*, *13*, 51–68. https://doi.org/10.1093/logcom/13.1.51

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). Chicago, IL: Open Court.

Hume, D. (Ed.). (1739/40). *A treatise of human nature. Book III: Of morals*. Available at http://www.gutenberg.org/ebooks/4705. Accessed November 11, 2018.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*, 169–188. https://doi.org/10.1017/s0140525x10003134

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

Kahneman, D. (2011). *Thinking, fast and slow*. London, UK: Macmillan.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*, *3*, 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Kant, I. (1998). *Critique of pure reason* (P. Guyer & A. Wood, Eds. & Trans.). New York, NY: Cambridge University Press. (Original work published 1781)

Kant, I. (2012). *Groundwork of the metaphysics of morals* (M. Gregor & J. Timmermann, Eds. & Trans.). Cambridge, UK: Cambridge University Press. (Original work published 1785)

Katsikopoulos, K. V., Schooler, L. J., & Hertwig, R. (2010). The robust beauty of ordinary information. *Psychological Review*, *117*, 1259–1266. https://doi.org/10.1037/a0020418

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, *48*, 507–531.

Knight, F. H. (1921). *Risk, uncertainty, and profit*. Boston, MA: Houghton Mifflin Harcourt.

Kornblith, H. (2002). *Knowledge and its place in nature*. Oxford, UK: Oxford University Press.

Leitgeb, H., & Pettigrew, R. (2010a). An objective justification of Bayesianism I: Measuring inaccuracy. *Philosophy of Science*, *77*, 201–235. https://doi.org/10.1086/651317

Leitgeb, H., & Pettigrew, R. (2010b). An objective justification of Bayesianism II: The consequences of minimizing inaccuracy. *Philosophy of Science*, *77*, 236–272. https://doi.org/10.1086/651318

Lewis, D. (1986). A subjectivist's guide to objective chance. In D. Lewis (Ed.), *Philosophical papers* (Vol. II, pp. 263–293). New York, NY: Oxford University Press. (Original work published 1980)

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology*, *1*, 65–82. https://doi.org/10.1177/0959354391011005

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, *118*, 393–437. https://doi.org/10.1037/a0024143

Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 119–140). New York, NY: Oxford University Press.

Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.). London, UK: Oxford University Press. (Original work published 1863)

Moore, G. E. (1903). *Principia ethica*. Cambridge, UK: Cambridge University Press.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–631. https://doi.org/10.1037/0033-295X.101.4.608

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*, 349–357. https://doi.org/10.1016/s1364-6613(00)01699-5

Oaksford, M., & Chater, N. (Eds.). (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York, NY: Oxford University Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge University Press. https://dx.doi.org/10.1017/CBO9781139173933

Pearl, J. (1990). System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In R. Parikh (Ed.), *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge* (pp. 121–135). San Francisco, CA: Kaufmann.

Pearl, J. (2009). *Causality*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Pelletier, F. J., & Elio, R. (1997). What should default reasoning be, by default? *Computational Intelligence*, *13*, 165–187.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*, 29–46. https://dx.doi.org/10.1037/h0024722

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353. https://doi.org/10.1126/science.146.3642.347

Popper, K. (2005). *The logic of scientific discovery*. Abingdon, UK: Routledge. (Originally published 1959)

Psillos, S. (1999). *Scientific realism. How science tracks truth*. London, UK: Routledge.

Quine, W. V. (1960). *Word and object*. Cambridge, MA: MIT Press.

Ramsey, F. P. (1990). Truth and probability. In D. H. Mellor (Ed.), *F. P. Ramsey: Philosophical papers* (pp. 52–109). Cambridge, UK: Cambridge University Press. (Original published 1926)

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Ridley, M. (1993). *Evolution*. Oxford, United, UK: Blackwell Scientific.

Salmon, W. C. (1957). Should we attempt to justify induction? *Philosophical Studies*, *8*, 33–48.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Samuels, R., Stich, S., & Bishop, M. (2012). Ending the rationality wars. How to make disputes about human rationality disappear. In S. Stich (Ed.), *Collected papers, Volume 2: Knowledge, rationality, and morality, 1978–2010* (pp. 191–223). Oxford, UK: Oxford University Press.

Schurz, G. (1997). *The is–ought problem: An investigation in philosophical logic*. Dordrecht, the Netherlands: Kluwer Academic.

Schurz, G. (2013). *Philosophy of science: A unified approach*. New York, NY: Routledge.

Schurz, G. (2014). Cognitive success: Instrumental justifications of normative systems of reasoning. *Frontiers in Psychology*, *5*, 625. https://doi.org/10.3389/fpsyg.2014.00625

Schurz, G. (2017). No free lunch theorem, inductive skepticism, and the optimality of meta-induction. *Philosophy of Science*, *84*, 825–839. https://doi.org/10.1086/693929

Schurz, G. (2018). Optimality justifications: New foundations for foundation-oriented epistemology. *Synthese*, *195*, 3877–3897. https://doi.org/10.1007/s11229-017-1363-6

Schurz, G. (in press, 2019). *Hume's problem solved: The optimality of meta-induction*. Cambridge, MA: MIT Press.

Schurz, G., & Lambert, K. (1994). Outline of a theory of scientific understanding. *Synthese*, *101*, 65–120. https://doi.org/10.1007/BF01063969

Schurz, G., & Thorn, P. D. (2012). Reward versus risk in uncertain inference: Theorems and simulations. *Review of Symbolic Logic*, *5*, 574–612. https://doi.org/10.1017/s1755020312000184

Schurz, G., & Thorn, P. D. (2016). The revenge of ecological rationality: Strategy-selection by meta-induction within changing environments. *Minds and Machines*, *26*, 31–59. https://doi.org/10.1007/s11023-015-9369-7

Shier, D. (2000). Can human rationality be defended a priori? *Behavior and Philosophy*, *28*, 67–81.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1–19.

Simon, H. (1991). Cognitive architectures and rational analysis: Comment. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 25–40). Hillsdale, NJ: Erlbaum.

Spiliopoulos, L., & Hertwig, R. (in press). Strategic uncertainty and incomplete information: The Homo heuristicus does not fold. In R. Hertwig, T. J. Pleskac, T. Pachur & The Center for Adaptive Rationality, *Taming uncertainty*. Cambridge, MA: MIT Press.

Stich, S P. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge, MA: MIT Press.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285. https://doi.org/10.1126/science.1192788

Thorn, P. D., & Schurz, G. (2014). A utility based evaluation of logico-probabilistic systems. *Studia Logica*, *102*, 867–890. https://doi.org/10.1007/s11225-013-9526-z

Thorn, P. D., & Schurz, G. (2016). Qualitative probabilistic inference under varied entropy levels. *Journal of Applied Logic*, *19*, 87–101. https://doi.org/10.1016/j.jal.2016.05.004

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, *11*, 92–107. https://doi.org/10.1080/17470215908416296

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140. https://doi.org/10.1080/17470216008416717

Wason, P. C. (1969). Regression in reasoning? *British Journal of Psychology*, *60*, 471–480. https://doi.org/10.1111/j.2044-8295.1969.tb01221.x

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*, 1341–1390.

Wundt, W. (1973). *An introduction to psychology* (R. Pinter, Trans.). New York, NY: Arno Press. (Original work published 1912)