Rethinking Rationality

Emmanuel M. Pothos[1]

Timothy J. Pleskac[2]

Running head: rethinking rationality

Affiliations/ Correspondence: 1. Department of Psychology, City, University of London, London EC1R 0JD, UK; Emmanuel.pothos.1@city.ac.uk

2. Department of Psychology, University of Kansas, Lawrence, KS, 66045, USA; pleskac@ku.edu

Word count including abstract but excluding references: 5946

Abstract

We seek to understand rational decision making and if it exists whether finite (bounded) agents may be able to achieve its principles. This aim has been a singular objective throughout much of human science and philosophy, with early discussions identified since antiquity. More recently, there has been a thriving debate based on differing perspectives on rationality, including adaptive heuristics, Bayesian theory, quantum theory, resource rationality, and probabilistic language of thought. Are these perspectives on rationality mutually exclusive? Are they all needed? Do they undermine an aim to have rational standards in decision situations like politics, medicine, legal proceedings, and others, where there is an expectation and need for decision making as close to 'optimal' as possible? This special issue brings together representative contributions from the currently predominant views on rationality, with a view to evaluate progress on these and related questions.

1. Tentative steps towards approaching rationality

Few notions are as important to our self-conception as our capacity for decision making or reasoning that is superior to that which was possible in our less evolved ancestors. In the behavioral sciences, the extensive and hugely influential research tradition on rationality exactly concerns the foundations of this kind of superior thinking that is assumed for our species (Arkes, Gigerenzer, & Hertwig, 2016; Anderson, 1991; Chase, Hertwig, & Gigerenzer, 1998; Chater & Oaksford, 2000; Cosmides, 1989; Gigerenzer & Selten, 2002; Oaksford & Chater, 2009; Savage, 1954/ 1972; Shafir & LeBoeuf, 2002; Simon, 1986; Stanovich & West, 2000). This question of rationality is not without significant practical consequences. There are many instances of human life when we rely on an assumption that humans are able to reach rational decisions. For example, in medical diagnosis, the emphasis is on decisions which are as good as possible, rather than merely adequate or expedient (Bergus et al., 1998). Likewise, in legal decision making, ideally all incidental bias in decision making would be eliminated and the conclusions that a jury would reach are the best possible ones, given the available information (Hastie, 1994). These are just two domains of many (e.g., driving and piloting are two more) where we suspect many people would prefer an 'ideal' decision maker, if one existed and could be trusted. Arguably, this contrast between a normative and descriptive emphasis is sharper in the study of human decision making than in any other field of inquiry.

But, how do we define rationality? The term is deep-rooted in everyday cognition. The *Merriam-Webster* online dictionary offers a meaning of rationality as "having reason or understanding." The Cambridge online dictionary defines the term as "the quality of being based on clear thought and reason, or of making decisions based on clear thought and reason." As a final example, the Oxford Learner's online dictionary offers us the definition as "the fact of being based on reason rather than emotions." In all cases, there is a theme of thoughtfulness and careful consideration, as opposed to impulsive or emotional decision making. Such a distinction is reminiscent of dual-process approaches to decision making (Elqayam & Evans, 2013; Kahneman, 2001; Sloman, 1996). Yet, these dictionary definitions present difficulties as they tend to focus on the thought processes of people, which are unobservable, as opposed to defining rationality in terms of observable properties like behavior.

More promising might be a more formal approach, one where a set of self-evident assumptions or axioms are utilized to formulate a theory of rationality (e.g., Luce & Raiffa, 1957; von Neumann & Morgenstern, 1944). Such an approach provides criteria to compare behavior against, much like we can judge whether a mathematical expression like 2 + 2 = 4 is correct or not. Such a notion of rationality as correctness implies consistency with a mathematical framework that we choose to guide inference. This approach has been profoundly influential particularly in areas like behavioral economics (Camerer, Loewenstein, & Rabin, 2004), game theory (Fudenberg & Tirole, 1991), decision analysis (Clemen & Reilly, 2013), and negotiation (Raiffa, 2007). However, it may be that consistency itself is insufficient for rationality and instead some additional arguments are needed as to *why* a particular mathematical framework provides a suitable normative standard (Arkes et al., 2016; Elqayam & Evans, 2011). For instance, the empirical evidence that incoherence—or at least the incoherence that people actually demonstrate—is actually costly is lacking (Arkes et al., 2016). In this case, what might be needed is an assessment of how well adopting a specific approach will result in achieving a particular goal (Anderson, 1990; Griffiths et al., 2015; Marr, 1982; Oaksford & Chater, 2009). But, a focus on goals brings to light factors that enter into the ability to achieve those goals be they behavioral, cognitive, or environmental. Indeed, one way to understand how we have begun to rethink rationality is to take seriously the role these factors play in our conceptualization of rationality.

Thinking of rationality in terms of goals and everyday cognition leads us to an alternative idea, that it might be possible to define rationality as human adaptiveness. But humans are not the only animals (and organisms) which appear to demonstrate adaptive qualities. Therefore, in this vein, we might be tempted to try to understand rationality via comparisons to other species (e.g., Santos & Rosati, 2015). But, these comparisons are sometimes not as straightforward as we might like, at least for the purpose of reaching clear definitions of rationality. Consider for instance the comparison between animals and humans on the ultimatum game (Güth et al., 1982). During the ultimatum game two parties interact anonymously and only once. The first player proposes how to divide a sum of resources with the second party. If the second player rejects this division, neither gets anything. If the second accepts, the first party gets its demand and the second gets the rest. A nice aspect of the ultimatum game is that the prediction of a rational decision maker who is self-interested and maximizes payoff is straightforward: the first party should offer as little as possible without offering nothing to the second party while the second party should accept any offer that is greater than zero. We know in some cases with the ultimatum game chimpanzees sometimes behave in exactly this self-interested, payoff maximizing way (Jensen et al., 2007; though see Proctor et al., 2013). At the same time, humans sometimes depart from the self-interested criterion, factoring in fairness in how they decide during the ultimatum game (Güth et al., 1990; see also Henrich et al., 2005). Thus, in the ultimatum game we are confronted with a possible disconnect where humans sometimes do not live up to what some might claim to be an ideal standard, while some animals do.

In other cases, the comparison between humans and other species confronts us with the possibility that both humans and other animals cannot live up to the abilities that we think are part of a repertoire of rational abilities. For instance, many people agree that one aspect of rational decision making is description invariance such that preferences for options should not depend on how they are described (Arrow, 1982). However, people regularly commit violations of description invariance falling prey to, for example, framing effects (Tversky & Kahneman, 1986). For example, people assess a hypothetical gamble as more favourable when it is described positively in terms of the likelihood of winning as opposed to describing it negatively in terms of the likelihood of losing (Levin et al., 1998). And, as it turns out, animals, like capuchin monkeys (Chen et al., 2006), bonobos, and chimpanzees (Krupenye et al. 2015), also seem to do something similar. In these experiments, the animals were found to prefer to trade a token with an experimenter who offered a smaller amount of food but sometimes augmented it (a gain) versus trading with an experimenter who initially offered more but sometimes reduced it (a loss). This preference was despite the fact that both experimenters offered the same expected payoffs. Similar parallels exist in other areas of decision making like the endowment effect (valuing an object more when possessing it than when seeking to acquire it) (Brosnan et al., 2007) and the certainty effect (the tendency to select the safer of two options when this option offers a good outcome with certainty; Shafir et al., 2008; Real, 1996). Perhaps these parallels disqualify these abilities as being part of a rational decision maker?

These examples illustrate three points complicating comparisons between humans and other animals. First, in some cases we do not have a precise notion of the ideal behavior. In the example concerning the ultimatum game, is self-interest the ideal behavior or is it behavior consistent with notions of fairness? And even if we accept one of those considerations, arguably each of those concepts are imprecisely defined. Second, as the description invariance example illustrates, we are (still) unsure of what are the critical competencies for rational thought. Third, there is a lack of clarity concerning what we should expect concerning the rationality of humans vs. other animals. On the one hand, we have an expectation that we—humans—are more rational in some qualitative way, than other animals, because of science, culture, etc. On the other hand, for non-human animal cognition, non-optimal behaviors might translate to survival disadvantages resulting in extinction. Note, these issues complicate an expectation that we might be able to understand rationality in humans, through comparisons with animal behavior, not the notion of approaching rationality as adaptiveness in general.

This special issue reviews some of these potential factors that seem to shape rational behavior. Here we offer three observations from these contributions, as well as theories of rationality more generally. First, we can distinguish between two broad categories of theories of rationality. On the one hand, there are theories which have put more emphasis on consistency with a particular standard, mathematical framework or set of principles. In this category we can include the perspectives to rational thought based on classical logic (Wason & Johnson-Laird, 1972), probability theory (Oaksford & Chater, 2009; Busemeyer & Bruza, 2011; Haven & Khrennikov, 2013), and expected and subjective expected utility theory (von Neumann & Morgenstern, 1944; Savage, 1954). On the other hand, there are theories and approaches which focus on how well one is able to achieve particular goals, such as theories of ecological rationality (Anderson, 1990; Brunswik, 1955; Gigerenzer et al., 1999; Hammond, 2000; Hertwig et al., 2019; Simon, 1990). Ultimately, we would like our chosen theories under these two categories to converge or at least find their functional equivalences, resulting in theory integration (Gigerenzer, 2017).

Second, there is extensive debate regarding the extent to which computational constraints ought to impact on rationality and how (Gigerenzer, 2001; Griffiths et al., 2015; Lewis et al., 2014; Lieder & Griffiths, 2019; in the Special Issue: Mohnert et al.; Hardy et al.; Hertwig et al.). Obviously, computational constraints help determine the possible behaviors that an agent (human or otherwise) can execute. But should computational constraints actually affect what we consider to be rational behavior? Some of the papers in this special issue will confront this important question. But computational constraints are not the *only* factors with potential to shape rationality. For instance, many inferences take place in communication settings, rich in conversational and pragmatic implicatures (e.g., Dulany & Hilton, 1991; Grice, 1975). The richness of linguistic settings suggests that language itself may shape how we approach rationality in the first place (e.g., Goodman et al., 2015; in the Special Issue, Tessler et al.). This discussion also highlights a potential limitation to many current decision theories that touch on rationality: they tend to focus on the use of quantifiable information. However, some of the information relevant to a decision may not be quantifiable. One could argue that the ability to reason and act on vague and qualitative information is part of humans' evolved decision making ability. Theories of rationality should not be limited to the quantified world, but also include the unquantified one as well (in this Special Issue: Shiffrin).

Finally, the debate on rationality has developed in a way that moral decision making is often not considered. This has not always been the case. Consider Pascal's Wager, whereby Blaise Pascal famously argued that a rational person should live as though (the Christian) God exists and endeavor to believe in God. This is because the costs of living as though God exists, when she or he does not, are greatly outweighed by the infinite gains to be had and infinite losses to be avoided if she or he does exist. At face value, it may seem that consistency with a moral standard is no different than achievement of any other goal. But there also appear to be peculiarities in moral decision making which imply that a view of rationality incorporating moral decisions may not be straightforward (e.g., Gamez-Djokic & Molden, 2016; Kahane & Shackel, 2010). Currently, research on moral decision making appears to incorporate ideas and models from research on rationality only to a limited extent. The two research directions have drifted apart. This is unfortunate, especially given how, as a society, we are facing novel problems with deep moral issues, such as the design of automated vehicles which can make ideal decisions concerning risks (Bonnefon et al., 2016).

2. Is rationality a useful construct?

This question is the elephant in the room. It arises partly because, as we outlined earlier, defining rationality is tricky. It also arises because even when we think we have arrived at a definition, this seems to change (e.g., Daston, 1988; Gigerenzer, Switjink, Porter, Daston, Beatty, Krüger, 1989; Gigerenzer & Selten, 2001; Hacking, 1975). For instance, initially, starting with Blaise Pascal and Pierre Fermat in 1654, maximizing expected value was considered rational. But then, the idea of maximizing expected value came under scrutiny and ultimately there was a shift to the idea that people ought to be maximizing subjective value or a utility instead. This major shift was partly a result of Daniel Bernoulli's solution to the St. Petersburg paradox (Bernoulli, 1738/1954). In the St. Petersburg paradox, a person, Alice, considers a simple gambling scenario, whereby she can toss a fair coin to e.g. triple her bankroll, B, on heads but lose everything on tails. So, on heads, she can win $3*B$, but on tails she only loses B. Should Alice accept this gamble or walk away with her current money, B? An expected value approach tells us that she should play, since the expected value of the gamble is $½* 3*B-½*B>0$. The paradox is that if Alice plays an infinite series of such gambles, expected value still recommends playing, but of course common sense tells us that Alice is eventually to lose *B* (i.e., lose all her money). Bernoulli partly resolved the paradox by claiming people should seek to maximize expected utility and not expected value, because expected utility is based on a principle of diminishing returns (c.f., Pothos, Shiffrin, & Busemeyer, 2014).

The idea of maximizing utility was appealing to economists (Jevons, 1871; Knight, 1921; Mill, 1844). However, the approach became bogged down with the question of what exactly was the utility being maximized, inviting ex post accounts of rationality. Eventually a satisfactory solution to this problem arose via von Neumann and Morgenstern's (1944) axiomatic expected utility theory, that provided a set of ex ante principles, such that if preferences meet these principles, then one could conclude that preferences were maximizing utility. These axioms and Savage's (1954/1972) extension in the domain of uncertainty have become the benchmark standard for rationality (Pleskac, Wallsten, & Diederich, 2015; Savage, 1954/1972; Wheeler, 2017).

A similar reconceptualization is evident when considering decisions with low or no utility. Since antiquity, an influential approach to rationality has been based on classical logic. The principles of classical logic were thought to be the basis of a mental logic, which could be employed productively to deal with most reasoning problems (e.g., Braine et al., 1995). In fact, it is not hard to see how logic can serve as foundation for many artificial intelligence applications. For example, Brooks (1986) used logic to develop some early AI systems for controlling a mobile robot. First order logics have been applied in automating pilots for combat pilot simulations (Jones et al., 1998), computer games (Wintermute et al., 2007), and a host of other areas (Laird et al., 1987). Classical logic is a deductive framework, that is, decision outcomes can be certainly correct vs. certainly incorrect. And this goes hand in hand with a falsificationist bias in inference, since we can never prove an assertion to be correct, however much confirmation we have (e.g., the sun rises every day), but we can disprove an assertion as incorrect from the existence of a single counterexample. In a famous demonstration, Wason (1968) showed that naïve observers do not appear to do this. That is, when confronted with a simple reasoning puzzle, naïve observers do not make the choices with the greatest potential for certain conclusions via disconfirmation, but instead seek confirmation: Oaksford and Chater (1994) showed that, if Wason's (1968) task is approached as one of trying to minimize uncertainty, then participants' behavior can be readily explained. The point is that behavior which appeared irrational from the point of view of one framework (classical logic) could be considered rational from the point of view of an alternative framework that was expanded to include classical probability. Overall, especially more recently, classical probability theory has been hugely influential in cognitive modelling (e.g., Lake et al., 2015; Tenenbaum et al., 2011; Tessler & Goodman, 2019).

But, an approach to rationality based on classical probability theory is not itself inviolate. For example, Tversky and Kahneman (1983; Kahneman et al., 1982) reported some very surprising and persistent discrepancies between human judgment and (baseline) classical probability prescription, such that Prob(A&B)>Prob(B). In a famous example, participants were told about a hypothetical person, Linda, very much described as a feminist and not at all as a bank teller. They were then asked to rank order the probability of different statements about Linda. Perhaps unsurprisingly, results revealed that participants considered more probable the statement that Linda is a feminist and bank teller, than just bank teller. This finding is called the conjunction fallacy and is one of the most famous findings in the decision literature (Moro, 2009). Such a judgment *appears* deeply problematic from a classical probability perspective, because classical probability is based on set theory and an intersection can never be more likely than a marginal (Tentori et al., 2004). Yet, Stephen J. Gould (1991) once said with the conjunction fallacy, “I know that the conjunction is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me—‘but she can’t be just a bank teller; read the description”’(p. 469; see also Gilboa, 2000). Bounded rational approaches (e.g., involving imperfect sampling, Sanborn & Chater, 2016; cf. Lieder & Griffiths, 2019) might offer a way to reconcile human judgment with classical prescription, but then we are faced with the problem of having to conclude that the unadulterated flavor of classical rationality is perhaps beyond the scope of any finite agent. Another way to reconcile such fallacies with rational expectation might be heuristics (Tversky & Kahneman, 1983). A challenge then would seem to have a clear rational framework for heuristics. It may be that the framework of ecological rationality is such an option (e.g., Gigerenzer et al., 1999; Hertwig et al., 2019). Yet another approach has been to argue that it is more rational to employ the alternative probability rules from quantum theory, when the questions are such that one question alters the meaning of the others (Busemeyer et al., 2011; Pothos et al., 2017).

Overall, the point is this: in the bigger picture it would seem that rationality—for better or worse—does not seem to be a static concept. As such, it is not clear that evaluating whether a behavior is consistent with a particular notion of rationality is helpful because what was rational yesterday may not be rational today and what is rational today may not be rational tomorrow. So, should we move on and focus our collective efforts on, say, understanding how people reason and decide? We say no. Rationality is a useful concept. We provide three arguments in support of its utility. First, a rational examination of a particular decision, as in the case of Wason's (1968) task, helps us isolate exactly why some people struggle with the response we (as experimenters) might be expecting. The appreciation that there are different rational perspectives (possibly applying under different circumstances), such that a decision may be unreasonable and incorrect from one perspective, but quite reasonable and correct from another, in itself is a noteworthy achievement and indeed requires a rational form of reasoning at a different level. This, we hope, illustrates our point that there is an inherent utility to be had by trying to understand rationality, as it would seem that people rely on it be they scientists, engineers, doctors, lawyers, mechanics, salespeople, pilots, and the list goes on, to reason and act effectively in this world.

Second, the concept of rationality is important because of the kind of information theories of rationality provide. Marr (1982) famously established three levels at which any information processing system must be understood. The computational level establishes the goal the system seeks to achieve and how it can be achieved functionally. The algorithmic level describes the processes by which the system works to meet this goal. The hardware level specifies how those processes are realized physically in the brain. From this perspective, theories of rationality primarily contribute to our understanding of the mind at the computational level, because they can provide a notion of the optimal way to achieve the stated goals and by doing so help isolate the problem space the mind is working in (Anderson, 1990; Griffiths et al., 2010; Chater & Oaksford, 1999).

A related reason is that an optimal/rational solution can, when available, serve as means of comparison between how the mind is solving a problem and how it ought to solve the problem. Consider the gaze heuristic, that is often used to illustrate the adaptive toolbox (e.g., Gigerenzer & Brighton, 2009), and which appears to be used by animals and humans alike to intercept flying objects like a fly ball, a frisbee, a thrown rock, or even flying prey (Hamlin, 2017; McLeod & Dienes, 1993; Shaffer Krauchunas, Eddy, McBeath, 2004). For example, in the case of a fly ball, how should a catcher behave, assuming she wants to intercept the ball and hopefully catch it? An effective strategy would be to first fixate on the ball and start running (after the ball is hit), keeping her angle of gaze constant (between 0º and 90º). By following this simple heuristic, a ball player can ensure they will intercept the ball. One can work out the optimal solution to the ball's trajectory based on the initial distance travelled, velocity, angle, air resistance, wind speed, spin etc (Hamilin, 2017; McLeod & Dienes, 1993). However, this gaze heuristic illustrates that a simple heuristic that ignores a lot of information can be quite effective (Gigerenzer & Gassmaier, 2011). The broader point is that knowledge of the optimal solution is quite useful to help make sense of psychological processes. It helps in identifying what information is being ignored or where the psychological approximation will fail and where it will succeed. The optimal solution may not always be available or even possible to find. Yet, when the optimal solution is known, it is certainly useful in understanding behavior. Arguably, this is part of what is achieved with the study of rationality.

Our third and final argument for the utility of rationality research is a motivational one: it is a moonshot. That is, it is akin to when John F. Kennedy, seemingly out of nowhere, set the apparently impossible goal of putting a man on the moon by the end of the decade. People did not believe it could be done. The technology was certainly not there, but neither was the human psyche. And yet this goal served to galvanize a great effort to create the necessary technology and overcome people's doubts of whether this could be ever accomplished. In an analogous way, the goal of understanding rationality is a moonshot. We may never achieve this goal, but our pursuit of this goal has certainly improved the human condition bringing us tools like logic, probability, utility, and others. Each of these tools has helped humans move from the dark and into, as John Locke (1690/1959) notes, the "twilight of reason" (Ch XIV). What more can we obtain? Perhaps by rethinking rationality we can start to understand how social, behavioral, cognitive, and environmental factors can help us further cut through the fog of uncertainty.

3. Outline of the special issue

The special issue brings together four recent perspectives on rationality, which illustrate some of the above ideas. The selection of the four approaches was in part based on subjective impression of the way the debate on rationality has been developing, the current state of the art, and the more innovative ideas with potential to transform our understanding of what it means to be rational. There is an emphasis here on the issue of rationality; the literature on decision making is vast, but there are plenty of models for which rationality is not covered in a detailed way. Inevitably, the objective to create a special issue also meant that we cannot be comprehensive regarding the relevant ideas. The four broad approaches represented in the special issue are best considered as a sample of (to our estimation) prominent current broad theories of rationality, selected in a way to offer as wide coverage of the theoretical landscape as possible. Many of the authors of the papers in this special issue took part in a symposium on "extending rationality" in the 20th Annual Meeting of the Cognitive Science Society (Pothos et al., 2019).

First, the approach of ecological rationality takes the position that cognitive processes, including simple heuristics, are not per se rational or irrational, but that their success rests on their degree of fit to relevant environmental structures (Hertwig et al., 2019). The key is therefore to understand how cognitive and environmental structures slot together and Hertwig et al. (this issue) makes the case that from this perspective a successful theory of rationality will need to account for three components: the mind's information processing, the environment to which the mind adapts, and the intersection between the environment and the mind. Ecological rationality emphasizes adaptative success and the goals of the organism (see also, Anderson, 1990, 1991; Payne, Bettman, & Johnson, 1993).

Consistency with the principles of a formal framework has been a frequent pathway to conceptualize rationality. As we discussed, one set of principles for conceptualizing rationality is logic and another classical probability theory. More recently researchers have been employing quantum theory as an alternative set of principles. Quantum theory offers a set of probabilistic rules for inference alternative to those of classical probability theory. Quantum cognitive models appear to do well when it appears that earlier questions create unique contexts or perspectives for subsequent ones, so that it is as if there are 'separate' probability spaces, instead of a single, all-inclusive one (Busemeyer et al., 2011; Pothos et al., 2017, 2021). Otherwise, such models resemble in many ways classical probability ones and in this special issue we offer two examples (Busemeyer & Wang, this issue; Wojciechowski et al., this issue).

Another approach to rationality has been to consider how resource or environmental constraints should impact on considerations of what is rational behavior. Resource-rationality classical probability models balance the requirement of consistency with classical probability theory with the computational cost of the corresponding operations. The closure requirement of classical probability theory and its implication that in most practical situations the resulting probability information would be intractably complex have been well-known in the decision community for a long time (indeed, arguably since Simon's, 1955, seminal work). However, how to exactly modify inference in a bounded-rational way is not straightforward and the resource-rational research programme is a promising way of doing so – in the present special issue we include two examples from this research programme (Mohnert et al., this issue; Hardy et al., this issue).

Finally, rationality clearly depends not just on the computation but also on the information computations take place on. Human communication is extremely rich in terms of the additional assumptions even very simple utterances appear to entail (e.g., Dulany & Hilton, 1991; Grice, 1975). The paper on probabilistic language of thought is an exciting attempt to offer a principled, formal model for how pragmatics interface with probabilistic information to support inference (Tessler et al., this issue).

Overall, at a broad and approximate level, one can say that both resource rationality (Mohnert et al., this issue; Hardy et al., this issue) and probabilistic language of thought (Tessler et al., this issue) are Bayesian approaches, but informed by different constraints regarding inference – in one case the emphasis is on the resource capacity of the agent and in the other case on the pragmatics of language communication. Quantum theory (Busemeyer & Wang, this issue; Wojciechowski et al., this issue) can be considered a Bayesian approach, but in a limited way (which may arise from resource constraints), so that in some cases non-classical effects would arise (such as the conjunction fallacy). Ecological rationality (Hertwig et al. in this Special Issue) is focused on goals and rationality concerns whether these goals are achieved, not on the inference process as such.

These four perspectives on rationality concern individual decision making. How do individual degrees of rationality translate to rationality at the societal or group level? Hahn (this issue) will address this problem. Also, many of the approaches to rationality we consider imply quantitative information about uncertainty. Is this a reasonable assumption in *all* cases? That the answer is yes seems like a fairly general assumption. Shiffrin (this issue) will question it.

4. Looking to the future

We end our introduction with a list of outstanding questions. Some of these questions might appear moonshots—too ambitious to be resolved in the immediate future. Still, as argued, we think there is utility in aiming for them.

First, there are two ways in which there is a concern that we might end up with a notion of rationality that is, in some sense, too weak, bounded rationality and rationality as adaptation. Regarding the former, rationality decision researchers (certainly all represented in this special issue) share the belief that we should be looking to understand bounded rationality, rather than rationality as such. But, is there a danger that this sets the rationality bar too low? A critic may argue that any decision might appear bounded rational, if the 'bounds' are just too low. Our response would be that what theories of bounded rationality accomplish is to offer a graded understanding of rationality: they allow us to identify decisions *more* rational than others, depending on realistic constraints. Relatedly, even if we are currently unable to offer an understanding of what is an ideal form of rationality, that is, rationality for agents with infinite resources, we can generally compare two decisions and decide which one is more rational.

Second, is there a danger that understanding rationality as adaptiveness (also) sets the bar too low (Anderson, 1990)? Understanding rationality as adaptiveness may mean rationality is not limited to human adaptiveness, but can be used to make sense of very simple agents or organisms, like fruit flies, in the same way as for complex organisms like our humble selves. So, there is a danger that we will end up with senses of rationality which are trivial, not because the capacity for rational behavior or function is bounded (as just above), but because the goals/ environment of the agent are too simple (in the extreme, a doorknob might be perfectly rational, in terms of its high adaption to the problem it is designed to address). The answer we offer is analogous to the one just above: adaptiveness is probably better understood as a graded notion, in terms of the complexity of the problem: more complex problems would require more elaborate forms of adaptation and so lead to more interesting notions of rationality.

Third, what about doctors, lawyers, jury members, engineers, pilots, or drivers? How can we guide them to be as close to ideally rational as possible? Our expectations for rationality should be informed by environmental considerations, which may impact not just on how achievable different levels of rationality are (this is the old-fashioned view of rationality), but also on what are appropriate goals and how the relevant environment impacts on these goals (Guess, 2004). That is, a complete theory of rationality should include conditions when and where a particular strategy is effective. For example, if we accept the premise that heuristics exploit structures of the environment, then we need to be able to predict when those structures will arise, so as to predict when the heuristics are effective. It follows that we need a theory of the environment. This is difficult and psychologists typically do not look at decision making this way. However, a theory of the environment seems a prerequisite before we can start delineating the scope of applicability of different approaches to rationality (Pleskac et al., 2021).

Fourth, should we expect a unique flavor of rationality? After all, in this special issue there are four separate approaches, all of which relate to each other, but also all of which have some unique predictions. It is somewhat unlikely that it will turn out to be the case that one approach to rationality (e.g., one of the ones considered presently) will be just correct and the others wrong, not least because any perspective with any traction in the literature would be supported by extensive theoretical development and empirical support. It is more likely that the eventual theory of rationality will be an agglomeration of components from the current theories, augmented with the range of factors (including resources, environment, pragmatics) which impact on what is rational, under the precise circumstances of a decision. Such circumstances might include the environment and culture within which a decision is situated (Guess, 2004; Pleskac et al., 2021). Additionally, it is possible that an eventual theory of rationality will include components which take into account putative vagueness in some of the information that is relevant to our decisions (Shiffrin, this issue).

Fifth, if people are not in general rational, should we worry about the effectiveness of democratic forms of government? As Surowiecki (2005; see also Hahn, this Special Issue) argued, decisions which are only marginally correct can be aggregated to improve the collective outcome. Strictly from the point of view of rational decision making, it may be tempting to entertain the possibility of confining decisions to the more rational members of a population (or even just the single most rational one?). At least as early as the Ancient Greeks, political scientists and philosophers have been debating this issue. Does research on rationality alter our perspective on what is the best approach? The only point we wish to make is that a cleaner conception of what rational thinking is may endow more members of a society with the tools to separate out good vs. poor evidence and arguments.

Sixth, is rationality the same as intelligence? Probably not. Rationality as currently researched concerns the consistency of the outcome of mental computation with certain standards, whether these standards are adaptive heuristics or the results of calculations from a probability theory. In fact, most theories of intelligence reference a range of competencies and abilities, *other than* consistency with a rational standard. Such abilities invariably concern, for example, speed or efficiency of processing (indeed, since the early work of Galton; see Jensen, 1982, or Deary et al., 2001). Even in theories which include, for instance, analytical reasoning (e.g., Gardner, 1987; Sternberg, 1985), this tends to relate to an ability to work through particular types of problems, less so to an ability to distinguish between rational vs. not rational lines of thought (it might be inferred that this latter ability is a prerequisite for the former). Rightly or not, currently there appears to be a disconnect between theories of rationality and of intelligence.

Finally, we return to the comparative issue of human vs. non-human animal cognition and the expectation that there is something fundamentally distinguishing between ourselves and other living organisms on earth. Are we closer to understanding whether differences in rationality is part of what enables humans to have a unique capacity for, for example, moonshots, compared to other organisms? Unfortunately, not currently, but at the very least we hope this paper and the ones in the special issue help sharpen the pertinent questions and the more promising directions for future work.

Acknowledgments

EMP was supported by ONRG grant N62909-19-1-2000. TJP was supported by the U.S. NSF grant 2121122

References

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

Anderson, J. R. (1991). Is human cognition adaptive. Behavioral and Brain Sciences, 14, 471-517. https://doi.org/10.1017/S0140525X00070801

Arkes, H. R., Gigerenzer, G., & Hertwig, R. (2016). How bad is incoherence?. *Decision*, *3*(1), 20--39. https://doi.org/10.1037/dec0000043

Arrow, K. J. (1982). Risk perception in psychology and economics. *Economic Inquiry*, *20*, 1–9. https://doi.org/10.1111/j.1465-7295.1982.tb01138.x

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

Bergus, G.R., Chapman, G.B., Levy, B.T., Ely, J.W., & Oppliger, R.A. (1998). Clinical diagnosis and order of information. *Medical Decision Making*, 18, 412–417. https://doi.org/ 10.1177/0272989X9801800409

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573-1576. https://doi.org/ 10.1126/science.aaf2654

Braine, M.D.S., O'Brien, D.P., Noveck, I.A., Samuels, M.C., Lea, B.L., Fisch, S.M., Yang Y. (1995). Predicting Intermediate and Multiple Conclusions in Propositional Logic Inference Problems: Further Evidence for a Mental Logic. *Journal of Experimental Psychology: General*, 124, 263-292. https://doi.org/10.1037/0096-3445.124.3.263

Brooks, R.A. (1986). A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation, 2, 14-23. https://doi.org/ 10.1109/JRA.1986.1087032

Brosnan, S. F., Jones, O. D., Lambeth, S. P., Mareno, M. C., Richardson, A. S., & Schapiro, S. J. (2007). Endowment effects in chimpanzees. *Current Biology*, *17*(19), 1704-1707. https://doi.org/10.1016/j.cub.2007.08.059

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193-217. https://doi.org/10.1037/h0047470

Busemeyer, J. R. & Bruza, P. (2011). Quantum models of cognition and decision making. Cambridge University Press: Cambridge, UK

Busemeyer, J. R., Pothos, E. & Franco, R., Trueblood, J. S. (2011) A quantum theoretical explanation for probability judgment 'errors'. *Psychological Review*, 118, 193-218. https://doi.org/10.1037/a0022542

Cambridge Dictionary. (n.d.). rationality. https://dictionary.cambridge.org/dictionary/english/rationality.

Camerer, C. F., Loewenstein, G., & Rabin, M. (Eds.). (2004). Advances in behavioral economics. Princeton University Press.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, 2(6), 206-214. https://doi.org/10.1016/S1364-6613(98)01179-6

Chater, N., & Oaksford, M. (2000). The rational analysis of mind and behavior. *Synthese*, 122, 93-131. https://doi.org/10.1023/A:1005272027245

Chen, M. K., Lakshminarayanan, V., & Santos, L. R. (2006). How basic are behavioral biases? Evidence from capuchin monkey trading behavior. *Journal of Political Economy*, *114*(3), 517-537.

Clemen, R. T., & Reilly, T. (2013). Making hard decisions with DecisionTools. Cengage Learning: Mason, OH https://doi.org/10.1086/503550

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

Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence*, 29, 389–399. https://doi.org/10.1016/S0160-2896(01)00062-9

Dulany, D.E., & Hilton, D. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, 9, 85-110. https://doi.org/10.1521/soco.1991.9.1.85

Elqayam, S., & Evans, J.S.B.T. (2011). Subtracting “ought” from “is”: Descriptivism versus normativism in the study of the human thinking. *Behavioral and Brain Sciences*, 34, 233-248. https://doi.org/10.1017/S0140525X11000483

Fudenberg, D., & Tirole, J. (1991). Game Theory. MIT Press

Elqayam, S., & Evans, J.S.B.T. (2013). Rationality in the new paradigm: Strict versus soft Bayesian approaches. *Thinking and Reasoning*, 19, 453-470. https://doi.org/10.1080/13546783.2013.834268

Gamez-Djokic, M. & Molden, D. (2016). Beyond affective influences on deontological moral judgment. *Personality and Social Psychology Bulletin*, 42, 11, 1522 – 1537. https://doi.org/10.1177/0146167216665094

Gardner, H. (1987). The theory of multiple intelligence. Annals Of Dyslexia, 37, 19-35. https://www.jstor.org/stable/23769277

Gigerenzer, G., & Selten, R. (Eds.). (2002). Bounded rationality: The adaptive toolbox. MIT press.

Gigerenzer, G. (2001). The adaptive toolbox. In G. Gigerenzer & R. Selten (Eds.), *Bounded Rationality* (pp. 37-50). Cambridge, MA: Massachusetts Institute of Technology.

Gigerenzer, G. (2017). A theory integration program. Decision, 4, 133-145. https://doi.org/10.1037/dec0000082

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

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

Gilboa, I. (2000). Theory of decision under uncertainty. Cambridge University Press: Cambridge, UK.

Goodman, N.D., Tenenbaum, J.B., & Gerstenberg, T. (2015). Concepts in probabilistic language of thought. In Eds. E. Margolis & S. Laurence, New Directions in the Study of Concepts, pp.623-653. MIT Press: Cambridge.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), Syntax and semantics 3: Speech acts (pp. 41-58). San Diego, CA: Academic Press.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357-364. https://doi.org/10.1016/j.tics.2010.05.004

Griffiths, T.L., Lieder, F., & Goodman, N.D. (2015). Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7, 217-229. https://doi.org/10.1111/tops.12142

Guess, C. D. (2004). Decision making in individualistic and collectivistic cultures. Online Readings in Psychology and Culture, 4, 1-18.

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

Güth, W., & Tietz, R. (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, *11*(3), 417-449. https://doi.org/10.1016/0167-4870(90)90021-Z

Hamlin, R. P. (2017). “The gaze heuristic:” Biography of an adaptively rational decision process. *Topics in Cognitive Science*, *9*(2), 264-288. https://doi.org/ 10.1111/tops.12253

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

Hastie, R. (Ed.). (1994). Inside the juror: The psychology of juror decision making. Cambridge University Press.

Haven, E. & Khrennikov, A. (2013). Quantum Social Science. Cambridge: Cambridge University Press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., ... & Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain sciences*, *28*(6), 795-815. https://doi.org/10.1017/S0140525X05000142

Hertwig, R., Pleskac, T. J., Pachur, T., & Center for Adaptive Rationality. (2019). Taming Uncertainty. Cambridge, MA, MIT Press. https://doi.org/10.7551/mitpress/11114.001.0001

Jensen, A. R. (1982). Reaction time and psychometric g. In *A model for intelligence* (pp. 93-132). Springer, Berlin, Heidelberg.

Jensen, K., Call, J., & Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science*, 318, 107-109. https://doi.org/10.1126/science.1145850

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

Jones, R., Laird, J., & Nielsen, P.E. (1998). Automated intelligent pilots for combat flight simulation. In AAAI-98, pp.1047-54.

Kahane, G., & N. Shackel. (2010). Methodological issues in the neuroscience of moral judgement. *Mind & Language*, 25, 561–582. https://doi.org/10.1111/j.1468-0017.2010.01401.x

Kahneman, D. (2001). Thinking fast and slow. Penguin: London, UK.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

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

Krupenye, C., Rosati, A. G., & Hare, B. (2015). Bonobos and chimpanzees exhibit human-like framing effects. Biology letters, 11, 20140527. https://doi.org/10.1098/rsbl.2014.0527

Laird, J., Newell, A., & Rosenbloom, P.S. (1987). SOAR: an architecture for general intelligence. AIJ, 33, 1-64. https://doi.org/10.1016/0004-3702(87)90050-6

Lake, B.M., Salakhutdinov, R., & Tenenbaum, J.B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350, 1332-1338. https://doi.org/10.1126/science.aab3050

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149-188. https://doi.org/10.1006/obhd.1998.2804

Lewis, R.L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, 6, 279–311. https://doi.org/10.1111/tops.12086

Lieder, F. & Griffiths, T.L. (2019). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43, 1-85. https://doi.org/10.1017/S0140525X1900061X

Locke, J. (1690/1959). *An essay concerning human understanding*. Dover.

Luce, R. D., & Raiffa, H. (1957). *Games and Decisions*. New York, NY: Dover Publications, Inc.

Marr, D. (1982). Vision. MIT Press: Cambridge, MA

Merriam-Webster. (n.d.). Rationality. Merriam-Webster. https://www.merriam-webster.com/dictionary/rationality.

McLeod, P., & Dienes, Z. (1993). Running to catch the ball. *Nature*. 362 https://doi.org/10.1038/362023a0.

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

Moro, R. (2009). On the nature of the conjunction fallacy. *Synthese*, 171, 1-24. https://doi.org/10.1007/s11229-008-9377-8

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. (2009). Précis of Bayesian rationality: the probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, 32, 69-120. https://doi.org/10.1017/S0140525X09000284

Oxford Learner's Dictionaries (n.d.) https://www.oxfordlearnersdictionaries.com/definition/english/rationality.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York, NY: Cambridge University Press.

Pleskac, T. J., Conradt, L., Leuker, C., & Hertwig, R. (2021). The ecology of competition: A theory of risk–reward environments in adaptive decision making. *Psychological Review, 128*(2), 315–335 https://doi.org/10.1037/rev0000261

Pleskac, T. J., Diederich, A., & Wallsten, T. S. (2015). Models of decision making under risk and uncertainty, In The oxford handbook of computational and mathematical psychology. New York, NY, Oxford University Press.

Pothos, E. M., Shiffrin, R. M., & Busemeyer, J. R. (2014). The dynamics of decision making when probabilities are vaguely specified. *Journal of Mathematical Psychology*, 59, 6-17. https://doi.org/10.1016/j.jmp.2013.09.001

Pothos, E. M., Busemeyer, J. R., Shiffrin, R. M., & Yearsley, J. M. (2017). The rational status of quantum cognition. *Journal of Experimental Psychology: General*, 146, 968-987. https://doi.org/10.1037/xge0000312

Pothos, E. M., Lewandowsky, S., Basieva, I., Barque-Duran, A., Tapper, K., & Khrennikov, A. (2021). Information overload for (bounded) rational agents. *Proceedings of the Royal Society B*, 288, 20202957. https://doi.org/10.1098/rspb.2020.2957

Pothos, E. M., Busemeyer, J. R., Pleskac, T., Yearsley, J. M., Tenenbaum, J. B., Goodman, N. D., Tessler, M. H., Griffiths, T. L., Lieder, F., Hertwig, R., Pachur, T., Leuker, C., & Shiffrin, R. M. (2019). Symposium on extending rationality. In Proceedings of the 41st Annual Conference of the Cognitive Science Society. Montreal, Canada: Cognitive Science Society.

Proctor, D., Williamson, R.A., de Waal, F.B.M., & Brosnan, S.F. (2013). Chimpanzees play the ultimatum game. *Proceedings of the National Academy of Sciences*, 110, 2070-2075. https://doi.org/ 10.1126/science.1145850

Raiffa, H. (2007). Negotiation analysis: The science and art of collaborative decision making. Harvard University Press: Cambridge, MA

Real, L. A. (1996). Paradox, performance, and the architecture of decision-making in animals. *American Zoologist*, *36*(4), 518-529. https://doi.org/10.1093/icb/36.4.518

Sanborn, A.N. & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20, 883-893. https://doi.org/10.1016/j.tics.2016.10.003

Santos, L. R., & Rosati, A. G. (2015). The evolutionary roots of human decision making. *Annual Review of Psychology*, 66, 321-347. https://doi.org/10.1146/annurev-psych-010814-015310

Savage, L. J. (1954/1972). The foundations of statistics (2nd ed.). New York: Dover Publications.

Shaffer, D. N., Krauchunas, S. M., Eddy, M., & McBeath, M. K. (2004). How dogs navigate to catch Frisbees. *Psychological Science*, 15, 437-441. https://doi.org/10.1111/j.0956-7976.2004.00698.x

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. Annual Review of Psychology, 53(1), 491-517. https://doi.org/10.1146/annurev.psych.53.100901.135213

Shafir, S., Reich, T., Tsur, E., Erev, I., & Lotem, A. (2008). Perceptual accuracy and conflicting effects of certainty on risk-taking behavior. *Nature, 453*, 917-920. https://doi.org/10.1038/nature06841

Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Simon, H.A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118. https://doi.org/10.2307/1884852

Simon, H. A. (1986). Rationality in psychology and economics. *Journal of Business*, S209-S224. https://www.jstor.org/stable/2352757

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1-19. https://doi.org/10.1146/annurev.ps.41.020190.000245

Stanovich, K. E. & West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? *Behavioral and Brain Sciences*, 23, 645–65. https://doi.org/10.1017/S0140525X00003435

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. CUP Archive.

Surowiecki, J. (2005). The Wisdom of Crowds. Anchor Books: New York.

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

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: a misunderstanding about conjunction? *Cognitive Science*, 28, 467-477. https://doi.org/10.1207/s15516709cog2803_8

Tessler, M.H. & Goodman, N.D. (2019). The language of generalization. *Psychological Review*, 126, 396-436. https://doi.org/10.1037/rev0000142

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjuctive fallacy in probability judgment. *Psychological Review*, 90, 293-315. https://doi.org/10.1037/0033-295X.90.4.293

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59*(4), S251-S278.

von Neumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press (1980).

Wason, P.C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*. 20, 273–281. https://doi.org/10.1080/14640746808400161

Wason, P.C. & Johnson-Laird, P.N. (1972). The psychology of reasoning: Structure and content. Harvard University Press.

Wheeler, G. (2018). Bounded rationality. In E. Zalta's (Ed) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/bounded-rationality/

Wintermute, S., Xu, J., & Laird, J. (2007). SORTS: A human-level approach to real-time strategy AI. In Proc. Third Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE-07)