Three paradoxes of bounded rationality

David Thorstad

Comments welcome, but please ask before citing

Abstract

My aim in this paper is to develop a unified solution to three paradoxes of bounded rationality. The first is the regress problem that incorporating cognitive bounds into models of rational decisionmaking generates a regress of higher-order decision problems. The second is the problem of rational irrationality: it sometimes seems rational for bounded agents to act irrationally on the basis of rational deliberation. The third is the rational inevitability of maximization: it seems that behavior must maximize some important quantity such as value or choiceworthiness in order to be rational, contradicting the claim that bounded rationality is a form of satisficing rather than maximization. I review two strategies which have been brought to bear on these problems: the way of weakening which responds by weakening rational norms, and the way of indirection which responds by letting the rationality of behavior be determined by the rationality of the deliberative processes which produced it. Then I propose and defend a third way to confront the paradoxes: the way of level separation.

1 Introduction

The study of bounded rationality presents new problems and paradoxes. My aim in this paper is to set out and study three of the best-known paradoxes of bounded rationality which I will term the problem of rational irrationality, the regress problem, and the rational inevitability of maximization.

These paradoxes are often studied separately. However, many of the same normative ideas have been brought to bear on each of the paradoxes. As a result, I will study the paradoxes together. Studying the paradoxes together avoids duplication of normative labor, gives a broader view of the normative strategies underlying solutions to the paradoxes, and allows us to assess these strategies globally by their track records across the

board. This discussion will point the way towards a new strategy which can be used to shed light on all three paradoxes.

Here is the plan. Section 2 presents three paradoxes of bounded rationality.[1] Sections 3-4 review two normative strategies that have been used to resolve the paradoxes. Section 5 proposes a new strategy, and Section 6 shows how this strategy can be used to confront the paradoxes.

2 Three paradoxes of bounded rationality

2.1 The problem of rational irrationality

It is widely agreed that bounded agents should sometimes decide how to act using a toolbox of decisionmaking heuristics (Gigerenzer and Selten 2001; Gigerenzer and Gaissmaier 2011). Heuristics typically use only some of the information available to agents or make only some of the inferences warranted by that information.

There are three reasons why rational decisionmaking is often heuristic. First, in many cases there is an accuracy-effort tradeoff between the quality of decisions and the cost of producing them (Johnson and Payne 1985). Heuristics often strike the best balance between decision quality and decision cost. Second, we sometimes lack the cognitive ability to carry out complex nonheuristic inferences, for example by fully calculating expected utilities or deducing complex theorems (Chickering 1996; Dagum and Luby 1993). We cannot be rationally required to use decisionmaking strategies that are beyond our abilities. Third, we face less-is-more effects: in some contexts heuristics reliably perform more complex decisionmaking procedures because heuristics are less prone to overfitting (Geman et al. 1992; Gigerenzer and Brighton 2009).

To illustrate the structure of heuristic decisionmaking, consider decisionmaking by

[1]Like many bounded rationality phenomena, these paradoxes arise both in the practical domain of rational decisionmaking and in the theoretical domain of rational belief formation. For brevity, I present each paradox in the special case of rational decisionmaking. When possible, I give examples designed to illustrate how the same paradoxes arise in rational belief formation.

satisficing (Selten 1998; Simon 1955). Imagine that you are selecting a hotel room for a conference. A satisficer first selects an aspiration level in one or more goods to guide her search. For example, the hotel room should cost less than $150, be within five miles of the conference, have at least three stars and offer continental breakfast. The satisficer then searches for options, in this case hotel rooms, one at a time until she finds an option which meets all of her aspirations. She then takes that option, in this case by booking the hotel room.

In many contexts satisficing is a rational decisionmaking heuristic, returning effective decisions at reasonable cost. The problem is that satisficing sometimes produces actions which look irrational. To see the problem, note three facts about satisficing. First, satisficing displays only a threshold sensitivity to quantity. Although hotel rooms are penalized for costing more than $150, satisficing does not discriminate between hotel rooms which cost $149, $100, or $1. Second, satisficing is insensitive to goods which are not reflected in the agent's aspirations. Our satisficer's decision is unaffected by the presence or absence of a hotel rewards program. Third, satisficing is noncompensatory: shortfalls along one dimension cannot be compensated by excesses along another. Our satisficer will not book a hotel without continental breakfast, even if the cost savings would be enough to buy the best breakfast in town.

Such is the price of cognitive efficiency. These vulnerabilities do not mean that satisficing is irrational. But they do mean that satisficing will sometimes produce actions which look irrational, even when by all appearance satisficing was a rational decisionmaking strategy and was applied in a rational manner. Because satisficing is noncompensatory, ignores relevant goods, and displays only a threshold sensitivity to quantity, satisficing can produce choices which are both counterpreferential and go against the balance of an agent's normative reasons. For example, our satisficer may pass up an excellent hotel room which costs $100, exceeds most of her aspirations and comes with a top-notch rewards program, but is six miles away, only to end up booking a barely-satisfactory hotel room at the top of her price range. This choice is likely counterpreferential: if we showed

both hotel rooms to the satisficer at the same time, she would make the opposite choice. And on most ways of filling out the story, the satisficer has passed up the room she has most reason to book in favor of a room she has less reason to book.

This example raises the *problem of rational irrationality*. On the basis of apparently rational decisionmaking procedures, agents can take actions which look irrational. Instances of the problem of rational irrationality abound. It has been argued that the poor are led to dramatically undersave and overborrow based on rational patterns of allocating scarce cognitive resources such as attention and computational bandwidth (Mani et al. 2013; Shah et al. 2012). And it is well known that even our most cherished heuristics can lead agents to take dominated actions (Birnbaum 2008), or ignore features of options such as their scope (Desvousges et al. 1992) and duration (Fredrickson and Kahneman 1993) which matter a great deal. In face of these problems, judgment and decisionmaking researchers have generally retained their insistence that decisionmaking heuristics can be rational ways for bounded agents to make up their minds.[2] But how can heuristics be rational if they lead to irrational actions?

Some theorists bite the bullet and insist that the problem of rational irrationality is no problem at all. Frank Knight held that:

> It is evident that the rational thing to do is to be irrational, where deliberation and estimation cost more than they are worth. (Knight 1921, p. 67).

On Knight's view, if acting rationally is more trouble than it is worth, than rationality requires us to act irrationally. This view carries an uncomfortable ring of paradox that most philosophers will want to avoid.

As a consequence, a good solution to the problem of rational irrationality should do three things. First, it should dispel the paradox by explaining why it cannot be rational to act irrationally, even as a result of rational decisionmaking. Second, it should explain why

[2]See for example Gilovich and Griffin (2002, p. 3) and Tversky and Kahneman (1974, p. 1130) in the heuristics and biases tradition; Lieder and Griffiths (forthcoming) in the Bayesian tradition; and Gigerenzer and Selten (2001) in the adaptive toolbox tradition.

we are nonetheless tempted to classify these cases as instances of rational irrationality. And third, it should help us to see how the many near-cousins of Knight's position in the contemporary empirical literature may be in an important sense correct.

2.2 The regress problem

Suppose you are deciding how much money to save for retirement this year. We might model the base-level decision problem that you face by letting the options available to you be numbers of dollars saved and building a probabilistic model of the effects of each savings strategy. Call this problem $\mathcal{D}_0$.

Bounded rationality theorists protest that $\mathcal{D}_0$ is an incomplete description of the decision problem that you face. Investment decisions are the outcome of costly and ability-constrained processes of deliberation. Instead of immediately making an investment, you may first choose to deliberate about your investment strategy using one of these processes. So we should enrich our decision problem by incorporating new options representing decision procedures over monetary investments. Call this enriched problem $\mathcal{D}_1$.

But decision procedures are not produced in a vacuum any more than the decisions that result from them. Instead of deliberating straightaway, you may engage in metacognitive strategy selection (Lieder and Griffiths 2017; Marewski and Schooler 2011) aimed at picking a good decision procedure over investment strategies. So we need to enrich the decision problem to a new problem $\mathcal{D}_2$, with an enlarged option space that includes metacognitive processes of strategy selection.

And now we seem to have talked ourselves into a muddle. Metacognitive strategies can also be chosen as a result of higher-order deliberation, so it looks as though we will have to enrich the decision problem to a new problem $\mathcal{D}_3$ with an option space incorporating these meta-metacognitive procedures. Proceeding in this way, we can generate an ascending hierarchy of decision problems which look to improve on their predecessors in exactly the same way at $\mathcal{D}_1$ improved on $\mathcal{D}_0$.

This line of reasoning has been taken to generate a family of *regress problems* for theorizing about bounded rationality.[3] To begin, it may be unclear if there is any complete description of the decision problem that bounded agents face. Many authors have resisted this way of taking the problem by holding that decision problems facing bounded agents can be described as a limit, for example the limit $\mathcal{D}_\omega$ of all the finite-level decision problems posed above, or even perhaps the limit $\mathcal{D}_\alpha$ of decision problems below some larger transfinite ordinal α (Lipman 1991; Mongin and Walliser 1988; Vassilakis 1992). But this raises a new threat: it may turn out that there is nothing which it is rational for bounded agents to do unless we can prove the existence of a solution to these limit problems which satisfies even minimal rationality constraints connecting the rationality of deliberation procedures to the rationality of their results.[4]

Another difficulty is that as we enrich our decision problems they come to look less and less like problems that bounded agents face. If we know anything about bounded rationality, it is that boundedly rational agents typically do not operate with rich problem descriptions full of higher-order facts about deliberative and metacognitive processes. This is because the cognitive cost of solving each problem $\mathcal{D}_i$ is a rapidly increasing function of i, and the cognitive benefits of increased complexity soon diminish. As we iterate problem-hierarchies into the transfinite, it becomes increasingly less plausible that the decision problems described have anything much to do with the problems that bounded agents face.

The upshot of this discussion is that a solution to the regress problem should say three things. First, it should say precisely which decision problem bounded agents face. Is it $\mathcal{D}_0$, $\mathcal{D}_1$, $\mathcal{D}_\omega$, or something else? Second, it should show that, at least in typical cases, there exists a rational action which can be taken in such decision problems. And third, a solution should show how the type of decision problem described can plausibly be said

[3]For discussion of what the regress problem could be, see Lin (2014) and Smith (1991).

[4]For example, Mongin (2000) finds that a solution exists only if the agent's deliberation procedure is a constant function, returning the same base-level action in each world state. Constant deliberation procedures are typically irrational because they are insensitive to evidence.

to be a problem that bounded agents confront.

2.3 The rational inevitability of maximization

It is often said that bounded agents satisfice rather than maximize. Many authors hold that it would be a significant conceptual and normative mistake to say that bounded rationality is a type of maximization or optimization (Elster 1983; Klein 2001). The trouble begins when we try to say how this can be true in an interesting way.

There are, of course, some uncontroversial senses in which failures to maximize can be rational. Consumer psychologists sometimes say that agents satisfice rather than maximize when agents are disposed to choose options that are reasonably high-quality, even if there is good reason to suspect that a better option exists. There is some evidence for the rationality of satisficing in this sense: those who are content with good enough are more satisfied, less anxious and regretful, and less likely to delay or avoid decisionmaking (Iyengar and Lepper 2000; Schwartz et al. 2002).[5] But for precisely this reason, consumer psychology does not give us a deep sense in which satisficing is a rational alternative to maximization. So understood, satisficing is rational because it maximizes satisfaction and minimizes anxiety, regret and decision avoidance. A similar case could be made for the rationality of satisficing in its traditional, heuristic sense introduced in Section 2.1.

The example of consumer psychology illustrates the seeming *rational inevitability of maximization*. This is the claim that rational behavior must maximize an important quantity such as happiness, value, choiceworthiness, or fulfillment of duty. Otherwise, the behavior would not be rational. If the rational inevitability of maximization is correct, then it is hard to see what sense there could be to the claim that bounded rationality is not a type of maximization. If bounded rationality did not involve maximizing what matters, then it would not be a type of rationality at all.

The rational inevitability of maximization may seem to be a narrow consequentialist principle. But in fact it is quite a bit broader than that. Many nonconsequentialists accept

[5]This is controversial. See Scheibehenne et al. (2010).

that behavior is rational just in case it maximizes value. They hold only that the most valuable actions are sometimes those which honor rather than promote value. Nonconsequentialists may also deny the grounding claim that actions are rational in virtue of being valuable. But both of these claims are compatible with the rational inevitability of maximization. More generally, the rational inevitability of maximization is compatible with the claim that rational behaviors maximize something other than value, such as choiceworthiness or fulfillment of duty. And even theorists who deny the rational inevitability of maximization altogether do not do so willy-nilly. They will still demand an explanation for how it can be rational to fail to maximize choiceworthiness or value.

As an illustration of the rational inevitability of maximization, consider bounded optimization models (Russell and Wefald 1991; Russell et al. 1995). Theorists in this tradition see bounded rationality as imposing constraints on the optimal solution to a decision problem. For example, certain actions are ruled out as beyond an agent's cognitive abilities, and costs of deliberation are incorporated. But given these constraints, the thought is that we can recast bounded rationality as an optimization problem in which agents should take the options that maximize value given problem constraints.

With the increasing plausibility and explanatory success of Bayesian bounded rationality models (Icard 2018; Lieder and Griffiths forthcoming), the idea that bounded rationality can be modeled as a type of bounded or constrained optimization has gained in popularity. But these models give a straightforward sense in which bounded rationality consists in maximizing expected utility. The challenge posed by the rational inevitability of maximization is therefore twofold. The first is to say whether and in what sense bounded rationality is a type of maximization. And the second is to say how, if at all, this claim is in tension with bounded optimization and similar maximizing approaches to the study of bounded rationality.

We have our marching orders. We want to see what bounded rationality theorists should say about the problem of rational irrationality, the regress problem, and the rational inevitability of maximization. In Sections 3-4, I review two strategies that are commonly

brought to bear on these problems. Then in Sections 5-6, I propose a third way to confront the paradoxes.

3 The way of indirection

Our first solution begins with the problem of rational irrationality. This problem arises because seemingly rational decisionmaking procedures such as satisficing can produce irrational-looking actions, such as passing up an excellent hotel room in favor of a barely satisfactory room. If we want to avoid the Knightian conclusion that it can be rational to act irrationally, a natural idea is to find some guarantee that the rational status of deliberation procedures cannot come apart from the rational status of the actions they produce.

The *way of indirection* does this by making the normative status of actions derivative on the normative status of deliberation procedures. The most common way to do this is to say that an action is rational if it results from a rational deliberation procedure. Then apparent cases of rational irrationality will turn out to be rational actions resulting from rational decision procedures. If it is rational to make up your mind by satisficing, then it is rational to book the hotel room that results.

The way of indirection is the orthodox solution to the problem of rational irrationality among bounded rationality theorists. Because important bounds such as the costs of computation are felt most strongly at the level of decision procedures, Herbert Simon held that the fundamental turn in the study of bounded rationality is the turn to a notion of procedural rationality which reflects the rational importance of decision procedures. Simon understood the notion of procedural rationality as the way of indirection suggests: "Behavior is procedurally rational when it is the outcome of appropriate deliberation" (Simon 1976, p. 66). Most theorists in the bounded tradition have followed Simon in holding that apparently irrational behavior is rational when it results from appropriate deliberation.

Some authors have extended the way of indirection to solve the regress and maximization problems by allowing indirect normative assessment to cover higher levels of deliberation such as metacognition. We can treat options in the base-level decision problem $\mathcal{D}_0$ as nondeliberative actions such as booking a hotel room, and options which first appear in the α + 1-st level $\mathcal{D}_{\alpha+1}$ as decision procedures over options at level α. This representation can be used to generate two types of indirect solutions to the regress problem.

The first solution class, *limit models*, proves that after a (typically uncountable) number of iterations the decision problems $\mathcal{D}_\alpha$ converge to a limit $\mathcal{D}^*$ which reflects all deliberative and nondeliberative options that agents face (Lipman 1991). The regress stops at this level. Limit models are used to claim that rational agents should take the optimal deliberative or non-deliberative option o^* in $\mathcal{D}^*$. Applying the way of indirection, it is held that lower-level deliberative and nondeliberative options are rational just in case they are entailed by o^*. For example, if o^* is decisionmaking by satisficing then it is also rational to book whichever hotel room results from satisficing and irrational to engage in any metacognitive or higher-order deliberation on the matter.

The second solution class, *convergence theorems* identifies a level α and a base-level option o. It is shown that for any $\beta > \alpha$, the rational option in $\mathcal{D}_\beta$ eventually leads the agent to do o (Mongin 2000). On this basis, it can be held that o must be the eventual result of rational deliberation. By indirection, this means that o is rational.

Both of these solutions to the regress problem entail and explain versions of the rational inevitability of maximization. Partisans of limit models hold that rationality requires taking the option in a fully-specified decision problem $\mathcal{D}^*$ which maximizes value. And partisans of convergence theorems hold that rationality requires taking whatever base-level option results in the limit of value-maximizing deliberation procedures. We also recover a sense in which bounded rationality is not merely a form of maximization. The rational base-level action o will often not be value-maximizing in the base-level decision problem $\mathcal{D}_0$.

How should we evaluate the way of indirection? There are many ways to push back here. One worry is that the way of indirection may count irrational-looking base-level actions as rational when they are produced by rational deliberation procedures. I press this case in Section 4. Another worry is that the way of indirection models agents as responding to long chains of decision problems or transfinite limits of these chains, which seems a substantial departure from the situation of bounded decisionmakers. I want to set these worries aside for now and focus on a challenge that gets to the heart of my disagreement with the way of indirection.

The challenge for the way of indirection is that it collapses normative structure. It takes rich examples full of many objects for normative assessment such as action, deliberation and metacognition, then forces us to assess each of these objects in the same way. This is admirably parsimonious, but sometimes it is too parsimonious to capture the normative data.

As an example, consider the Cognitive Reflection Test. The Cognitive Reflection Test asks questions such as the following:

> **(Bat and Ball Problem**) A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost? (Frederick 2005).

The Bat and Ball Problem triggers an initial, frugal reasoning process based on attribute substitution (Kahneman and Frederick 2002).[6] This process simplifies the problem statement by replacing '$1.00 more than' with '$1.00', and from there quickly reasoning to the conclusion that the ball costs 10 cents. Many highly intelligent people fail to override this initial reasoning process, outputting the incorrect judgment that the ball costs 10 cents.

Something irrational has happened here, but what does that irrationality const in? Assume for simplicity that reasoning processes go no higher than $\mathcal{D}_2$, the level of metacognition. This is the highest level about which we have good descriptive data in most contexts,

[6]The time course of heuristic and nonheuristic reasoning in these problems is a matter of controversy (Bago and De Neys 2017; Travers et al. 2016). To my knowledge, all of my arguments in this section can be reformulated to accommodate other leading views.

and the problems that I raise for the way of indirection will only get worse as we move to higher levels.

There is mounting evidence that the Bat and Ball Problem does not involve irrational metacognition.[7] We often simplify problem statements to speed our reasoning. The job of rational metacognition is not to always override or double-check our simplified forms of reasoning. We should only override heuristic reasoning when features of the reasoning problem signify that we are likely to make a mistake. The fact that we sometimes fail to metacognitively override fallacious reasoning processes is not a sign of irrational metacognition. The only way to prevent this from happening would be to always override heuristic reasoning, contradicting the considerations advanced in favor of heuristic reasoning in Section 2.1.

The point is that the numbers in the Bat and Ball Problem are carefully chosen to trick rational metacognitive processes.[8] Most nearby problems do not trigger unreliable heuristic reasoning, and short of explicit calculation agents have no way to tell that this problem is any different. As a result, many theorists have taken the Bat and Ball Problem to involve rational metacognition.

But then where is the irrationality in the Bat and Ball Problem? We would like to say that this problem involves irrational belief and deliberation on the basis of rational metacognition. Agents select an inappropriate form of deliberation, namely attribute substitution, and as a result form inappropriate beliefs, but they do so on the basis of rational metacognitive monitoring and control of deliberation. However, the indirect approach cannot say this. The indirect approach will have to say that there is nothing

[7]For evidence of rational metacognitive monitoring, compare 'conflict' tasks such as the Bat and Ball Problem to 'no-conflict' versions of the same tasks in which no risky heuristic process tempts reasoners. In conflict tasks agents take longer to produce answers and confidence judgments (Bonner and Newell 2010; De Neys et al. 2013; Johnson et al. 2016); report lower metacognitive feelings of rightness and increased feelings of error (Gangemi et al. 2015; Thompson et al. 2011; Thompson and Johnson 2014); and suggest sensitivity to conflict in process measures such as gaze tracking and talk-out-loud procedures (De Neys and Glumicic 2008). For evidence that metacognitive control has been exercised in conflict tasks see (De Neys and Glumicic 2008; Simon et al. 2015), and for evidence that this control is rational see (Jackson et al. 2016).

[8]To see how numbers matter, note that we do not make the same mistake when the numbers are changed, as in the Banana and Bagel Problem (Frederick 2005).

irrational in our response to the Bat and Ball Problem if that response results from rational metacognition. And here it looks as though the indirect view errs by refusing to divorce the normative status of deliberation and its outcomes from the normative status of the metacognitive processes that produced them.

Exactly the same structure is found in decisionmaking problems. Sometimes agents act irrationally and deliberate irrationally about how to act, but do so on the basis of rational metacognition. This would happen, for example, if I asked you to make a small wager about the cost of the ball. Here the way of indirection says that irrational-seeming actions and deliberations are rational when they result from rational metacognition. This claim removes the possibility of paradox by preventing us from saying what we would like to say: that the agent has acted and deliberated irrationally, but metacognized rationally. The way of indirection collapses normative structure by forcing us to make the same assessment of all three objects.

4 The way of weakening

Our second solution begins with the rational inevitability of maximization. The rational inevitability of maximization is the claim that actions must maximize some important quantity such as choiceworthiness or value, otherwise they would not be rational. In some applications the rational inevitability of maximization can seem overly demanding.

Suppose you are at the grocery store choosing among a shelf full of artisanal vinegars. It is very easy for you to make a satisfactory choice. Just pick any affordable vinegar with the right flavor profile and a few reasonable indicators of quality. But if choosing rationally requires making the best possible choice, then it looks like you are doomed to spend a half-hour buying vinegar or else run a substantial risk of acting wrongly.

Many authors have taken such examples to suggest that rational norms for bounded agents need to be weakened in order to make them less demanding. The *way of weakening* confronts the paradoxes by developing weakened forms of traditional rational norms

which are appropriate for bounded agents.

The best-known philosophical exemplar of the way of weakening is Michael Slote's satisficing consequentialism (Slote 1984).[9] Consequentialists traditionally apply a maximizing standard of correctness on which actions are right or rational just in case they are best.[10] Satisficing consequentialism weakens the traditional consequentialist view by applying a satisficing standard of correctness on which actions are right just in case they are sufficiently good, even if not the best.

Satisficing consequentialism yields an immediate solution to the rational inevitability of maximization by giving a principled reason to reject the datum. The rational inevitability of maximization assumes that standards of correctness are maximizing, whereas satisficing consequentialism holds that standards of correctness should be satisficing. It is right to buy any sufficiently good bottle of vinegar, even if a better bottle is available.

Satisficing consequentialism will also reduce the problem of rational irrationality. Many cases used to motivate the problem of rational irrationality involve satisfactory action on the basis of rational deliberation. For example, our satisficing hotel-shopper books a merely satisfactory hotel room on the basis of rational heuristic deliberation. In these cases, satisficing consequentialism will say that agents both act and deliberate rationally because their actions and deliberations are both satisfactory. As a result, many of the cases used to motivate the problem of rational irrationality will not involve any type of irrationality.

Some authors hold that satisficing solves the regress problem. The regress problem gets off the ground because it is hard to see how agents can act optimally without engag-

[9] Another prominent exemplar is Christopher Cherniak's minimal rationality program (Cherniak 1981, 1986). As always, each position deserves its own separate treatment, although many of the remarks made here will generalize to other programs. It is important to recall that Cherniak's primary motivation was the predictive-explanatory project of finding the minimal conditions for the aptness of describing agents using intentional vocabulary. That may be a separate and compatible project (Bermúdez 2009).

[10] Here I will not be concerned with the difference between rightness and rationality as normative categories. One way to justify this omission would be to adopt a reason-responsiveness view on which rationality and rightness are both determined by the totality of relevant normative reasons. On many versions of this view, what it is right to do and what we are rationally required to do will be the same thing (Lord 2017).

ing in higher-order deliberation. But it may be easier to imagine that agents could act satisfactorily well without further deliberation. Most famously, John Elster claims that the right way out of the regress problem is to decide, at some level $\mathcal{D}_n$ to deliberate no further and act (Elster 1983). Elster thinks that the right way to take this action will be through satisficing, and that this will be a fully rational thing to do.

One worry for the weakening approach is that it opens the door to irrational forms of submaximization.[11] Slote considers the example of a home-seller who lists her home for a price that she considers satisfactory. Slote asks us to imagine that the home-seller could fetch a higher price with no added uncertainty or delay. Asked why she does not seek more money for her house, the home-seller replies that she is a satisficer: for her, a satisfactory price is good enough. Satisficing consequentialism permits her to set any satisfactory price, even if more money could be obtained with no additional cost.

I think that any plausibility attaching to the satisficing consequentialist's verdict in this case rests on the difficulty of imagining what Slote asks us to imagine: a case in which raising the price would not increase the time-to-sale, decrease the likelihood of sale, or invite greedy relatives to ask for their share of the pie. If we set these assumptions aside, it is difficult to see how the satisficer can rationally turn down free money. Perhaps we are to imagine that the satisficer has no desire to spend more money on herself. But then she could take the money and donate it to charity. Or perhaps the satisficer doubts that she could spend the money more productively than the home-buyer. But then we no longer have a failure to maximize value. These difficulties have led many authors to think that satisficing consequentialism should be construed in some other way, for example as a point about virtuous dispositions rather than a revision to standards of correctness.[12]

Turn next to the problem of rational irrationality. This problem arises in many classic heuristics and biases tasks, where rational-seeming heuristics produce irrational-seeming judgments and decisions. One datum to explain in these cases is why participants, when

[11] On this point see Bradley (2006), Byron (1998), Mulgan (2001), and Pettit (1984) among others.

[12] On this point see Swanton (1993, 2004) and perhaps also Slote (2004).

confronted with their errors, often withdraw or repudiate them. If I tell you that you have borrowed too much money, booked the wrong hotel room, or judged a conjunction to be more probable than one of its conjuncts, you may retort that you did so as a result of rational deliberation. But you will certainly not say that your actions and beliefs are rational and hence do not need to be changed. However, insofar as satisficing consequentialism can make progress on the problem of rational irrationality in these tasks, it will do so by saying that the offending beliefs and actions are rational because they are satisfactorily good. This wrongly suggests that participants in many classic tasks should stick to their guns.

Another way of making this point is that many of the base-level actions used to motivate the problem of rational irrationality are themselves blameworthy. For example, return to the previously-cited claim that poverty traps such as overborrowing and undersaving can result from rational patterns of decisionmaking. Roughly, the idea is that agents have only so much time, computational bandwidth and attention to allocate to reasoning. The working poor are faced with an abundance of high-stakes, short-term challenges. As a result it can be rational for the poor to allocate most of their cognitive resources to evaluating the short-term consequences of their actions (Mani et al. 2013; Shah et al. 2012). However, this short-termist pattern of resource allocation produces substantial overborrowing, because short-term capital needs will be prioritized over long-term debt.

Suppose now that a mother borrows too much money because she is busy at work and occupied thinking about how to pay for food, educate her children, and keep the lights on. Tragically, she is unable to repay the loan and as a result her family is evicted. As much as we should stress the fact that she took the loan as a result of rational patterns of deliberation, we would also like to say that the mother is blameworthy for taking the loan. In particular, we know exactly who can blame her: her family. But if this is right, then she must have acted irrationally or else her actions would not be blameworthy.

On the regress problem, I confess that I cannot see how satisficing is supposed to

help. If the point is that decisions at some level should be made without any higher-level deliberation, then we are in agreement. But this is precisely the point that a solution to the regress problem must argue for. The argument cannot be that in acting without deliberation we are guaranteed to act satisfactorily, because we have no such guarantee. And the point cannot be that decisionmaking by satisficing itself can be done without further deliberation. In its original heuristic sense, satisficing requires the specification of an aspiration level, the ordering of options during search, and procedures for adjusting aspirations as a result of new evidence. All of these can be and frequently are selected on the basis of higher-order deliberation (Caplin and Martin 2011). In asking how to conduct this deliberation, we take up a higher-order decision problem and by all appearances, the regress continues.

So far, we have reviewed and rejected two traditional solutions to the paradoxes: the way of weakening, and the way of indirection. The lessons learned from these discussions point us towards a third way to solve the paradoxes.

5 The way of level separation

In this section, I describe a third way of approaching the paradoxes, the *way of level separation*. The way of level separation consists of three normative claims. The first two of these claims deny the characteristic claims made by the ways of indirection and weakening, in order to avoid their pitfalls.

First, normative assessment should be *direct*, not indirect. The rationality of action, deliberation, metacognition, and higher-order deliberation is determined by separate applications of standards of correctness at each of these levels. For example, direct consequentialism would say that an action such as booking a hotel room is rational just in case the expected utility of booking that room is at least as great as the expected utility of booking any other room (Kagan 2000; Pettit and Smith 2000). And similarly, it is rational to deliberate about booking a hotel room using the satisficing heuristic just in case the

expected utility of deliberating by satisficing is at least as great as the expected utility of deliberating in any other way. These assessments can come apart, for example because the cognitive costs of deliberation bear on the expected utility of decision procedures but not on the expected utility of hotel bookings.

Direct normative assessment enforces a sharp level-separation between higher- and lower-order decision problems. This means that unlike indirect normative theories, the way of level separation will not collapse normative structure. We can easily countenance cases in which rational metacognition leads to irrational first-level deliberation and irrational base-level actions. If rational metacognizers will not intervene in low-stakes decisionmaking without substantial evidence that they are deliberating irrationally, then rational metacognition will sometimes involve letting irrational deliberation continue. Because each of these normative claims applies at a separate level, there is no pressure to deny any of these claims by collapsing levels.

Second, standards of correctness should be *maximizing*, not satisficing. It is rational to do what is most valuable or choiceworthy, and irrational to do anything less. We can soften this claim in the usual ways. For example, we could hold that goodness or even rationality is a scalar property and that satisfactory actions may be very good or even highly rational, even if not the best or rational full-stop. But by refusing to weaken the standards of rationality, we also avoid the unpalatable normative consequences of satisficing views. We can say why it is irrational for a home-seller to turn down free money, whatever the amount: it would be better to take the money in order to benefit herself or others.

For the way of level separation to be acceptable as a theory of bounded rationality, we will also have to say something about the sense in which the way of level separation is a theory of procedural rationality. Here is something that Herbert Simon might easily have said about procedural rationality, and which I think he should have said.

A large part of normative theory asks questions about *substantive rationality*: the rationality of base-level actions such as borrowing money. But it is also important to ask

questions about *procedural rationality*: the rationality of decisionmaking procedures which produce base-level actions, or the rationality of higher-order decisionmaking procedures which guide lower-level decisionmaking procedures.[13] This is important because key cognitive bounds such as the cost of deliberation are felt most strongly at the procedural level.

Traditional normative theory is *outcome-focused*, asking questions primarily about substantive rationality and assuming as a default stance that these substantive questions will settle important normative questions in other areas. By contrast, the bounded tradition holds that normative theory should be *process-focused*, primarily asking and applying questions about procedural rationality.

This is not merely a change in emphasis. Consider once again the woman who overborrows as a result of rational patterns of attention allocation and as a result is evicted from her home. Now suppose you are a hard-line luck egalitarian. You think it is unacceptable for people to be deprived of homes as an outcome of anything less than their own irrational choices. But you think it is no injustice for people to wind up homeless as a result of their own choices.

Here you could take either an outcome-focused or a process-focused interpretation of irrational choice. On an outcome-focused interpretation, the woman has made an irrational choice, namely borrowing too much money. As a result, she has not been treated unfairly if she is left homeless.[14] On a process-focused interpretation, the woman has made a rational choice, namely allocating her attention as best she can to the many cognitive challenges that she faces. As a result, she will be treated unfairly unless she is housed. The third claim made by the way of level separation is that normative assessment should be process-focused. In most contexts, facts about procedural rationality deserve at least as much emphasis as facts about substantive rationality.

[13]The epistemological analogue of this claim is emphasized by work in the epistemology of inquiry. See Friedman (forthcoming) and Levi (1991, 2012).

[14]This is a version of the harshness objection (Voigt 2007) to luck egalitarianism. One of the upshots of my discussion is that a process-focused interpretation eases, but does not eliminate the harshness objection.

Summing up, the way of level separation advocates direct rather than indirect forms of normative assessment; maximizing rather than satisficing standards of correctness; and process-focused rather than outcome-focused normative theorizing. What light can the way of level separation shed on the paradoxes?

6 Solving three paradoxes of bounded rationality

6.1 The problem of rational irrationality

The problem of rational irrationality is that it sometimes seems rational for bounded agents to act irrationally on the basis of rational deliberation. We saw in Section 2 that a good solution to the problem of rational irrationality should do three things. First, it should dispel the problem by showing why it cannot be rational to act irrationally. Second, a good solution should explain why we were tempted to say that it can be rational to act irrationally. And third, a good solution should explain how neighboring statements made in the empirical literature can be true in a nonparadoxical sense.

On the first point, the way of level separation dissolves the problem of rational irrationality by applying standards of correctness directly at each level. Actions are rational if they meet these standards and irrational otherwise. The fact that an agent deliberated rationally cannot change the fact that she acted irrationally as a result.

On the second point, the way of level separation explains why we are tempted to talk of rational irrationality by emphasizing that normative assessment should be process focused. In an intellectual climate where most normative theory is outcome-focused, the claim that a bounded agent has acted irrationally sometimes obscures an important normative fact: that she did so as a result of rational deliberation. In our zeal to emphasize the fact that agents have deliberated rationally, we are tempted to go too far and claim that the actions which result from this deliberation are also rational. That is an understandable temptation which level separation trains us to resist.

On the third point, researchers in the heuristics and biases tradition often say in

one breath that typical heuristics are rational, but also that these heuristics violate normative standards. The introduction to a leading anthology combines these claims in near-Knightian fashion.

> Although ... heuristics are distinguished from normative reasoning processes by biased judgments, the heuristics themselves are sensible estimation procedures that are by no measure 'irrational'. (Gilovich and Griffin 2002, p. 3).

In what sense can heuristic decision procedures be rational if they also violate normative standards? Level separation provides the natural answer. The heuristics themselves are rational decision procedures. On occasion, the decisions that they produce violate rational norms on action. But this fact does not impugn the rationality of the heuristics themselves, because it does not undermine any of the grounds given for the rationality of heuristic decisionmaking in Section 2.1.

6.2 The regress problem

The regress problem is that bounded agents seem to be confronted with an infinite regress of higher-order decision problems. We saw in Section 2 that a good solution to the regress problem should do three things. First, it should say which of these decision problems a bounded agent actually faces. Second, it should reassure us that these problems will typically have at least one rational solution. And third, it should tell us how the type of decision problem described is something that a bounded agent could face.

On the first point, level separation holds that bounded agents are confronted with multiple decision problems: how to act, decide, metacognize and the like. More precisely, we face any of these problems in which there are options available to us in our capacity as rational agents. On many views, this means that we face only the first few levels of decision problems, either because we lack the capacity to act in response to higher-order considerations or else because we can only do so in a sense that is too subpersonal

and insufficiently volitional or reason-responsive to count as an act of rational agency. But some Bayesians may think that we agentially confront each problem in an infinite hierarchy, and the way of level separation does not rule this view out in principle.

On the second point, level separation ensures that the multiplicity of decision problems does nothing to interfere with the existence of rational solutions to each problem. Of course, a lack of solutions may arise due to standard sources of rational dilemmas. And the rational solution to some problems may be to do nothing at all, for example because any type of twelfth-order deliberation would be too costly and unreliable to be worth engaging in. But each decision problem that agents face is a perfectly well-posed decision problem and there is no need to find an equilibrium or limiting solution to all of these problems at once in order to reassure ourselves that the problems are individually soluble.

On the third point, level separation helps us to see how the decision problems we face are problems that bounded agents could confront. As I have said, bounded agents may only face the first few levels of decision problems because we are incapable of agentially confronting higher-order decision problems. This fact does not mean that the decision problems we face are ill-formed or lead to regress. And as we confront these decision problems, the way of level separation does not demand that agents represent the problems to ourselves in any complex way. For example, an increasingly popular view of metacognition holds that we solve $\mathcal{D}_2$ largely by attending to a class of *noetic feelings* produced by our experience of deliberation, rather than by forming and analyzing metacognitive beliefs about our own deliberative processes. It is not especially difficult to see how bounded agents could be capable of forming and responding to such feelings, and a good deal of work has been done to show how humans (Arango-Muñoz and Michaelian 2014; Proust 2013) and even perhaps nonhumans (Hampton 2009) can use these feelings to metacognize effectively.

6.3 The rational inevitability of maximization

The rational inevitability of maximization is the claim that behavior by bounded agents must maximize some important quantity such as value or choiceworthiness, or else it could not be rational. The rational inevitability of maximization puts pressure on the claim that it is a normative or conceptual mistake to view bounded rationality as a form of maximization. We saw in Section 2 that a good response to the rational inevitability of maximization should do two things. First, it should say in what sense boundedly rational behavior maximizes some important quantity, or why it can be rational to fail to maximize in any sense. And second, it should say how, if at all, the theory of bounded rationality comes apart from bounded optimization models.

On the first point, the way of level separation captures the rational inevitability of maximization in the natural way, as the claim that standards of correctness should be maximizing. Bounded rationality is a maximizing theory insofar as agents are rationally required to act as well as possible given their bounds. This means that we need to recover some other sense in which bounded rationality is not a form of maximization.

Many things can be said here, but the second point gives us one natural way to interpret the claim that bounded rationality is not a form of maximization, namely by showing how bounded rationality comes apart from traditional bounded optimization models. As we saw, bounded optimality models typically assume that agents behave optimally at some top level such as metacognition, and then say that rational behavior at lower levels is whatever results from rational behavior at the top level. Section 3 argued that this feature of bounded optimality models is a normative mistake. Facts about boundedly optimal deliberation at a high, formally constructed level may have little bearing on rational behavior at lower levels. This gives an important sense in which theories of bounded rationality make different normative predictions than traditional bounded optimization approaches.

The way of level separation elaborates on the rational inevitability of maximization by revealing what we might call the *rational tragedy of maximization*: as a direct causal

result of optimal deliberation, bounded agents may be led to act suboptimally. We can deliberate exactly as we ought and still end up with a mediocre hotel room or an eviction notice. And it may be that the only way to avoid these outcomes would be to deliberate suboptimally.

Traditional solutions to the paradoxes such as the ways of weakening and indirection are designed to lessen or eliminate the rational tragedy of maximization. But the rational tragedy of maximization is here to stay. However, the tragedy of maximization keeps good philosophical company. We have long known that rational rules, dispositions and motive sets can on occasion produce irrational behaviors (Adams 1976; Harsanyi 1977). It may be rational to love your family even if you will sometimes be led to act selfishly out of love (Parfit 1984; Railton 1984). And it may be rational to be disposed to punish norm violations even if this disposition sometimes leads you to engage in costly punishments that do more harm than good.

At the same time as rationality generates these tragedies, it also limits their scope. It is rational to love your family because even though your love will occasionally lead you to act irrationally, in most cases your love will have a salutary effect on decisionmaking. So too, when it is rational to make decisions heuristically this is because even though heuristic decision procedures occasionally produce irrational actions, they typically produce rational actions. Insofar as bounded rationality is process focused, it urges us to pay attention to the fundamental rationality of our decision procedures and shed fewer tears over the occasional irrationality of decisions that result.

References

Adams, Robert Merrihew. 1976. "Motive utilitarianism." *Journal of Philosophy* 73:467–481.

Arango-Muñoz, Santiago and Michaelian, Kourken. 2014. "Epistemic feelings and epistemic emotions." *Philosophical Inquiries* 2:97–122.

Bago, Bence and De Neys, Wim. 2017. "Fast logic? Examining the time course assumption of dual process theory." *Cognition* 158:607–617.

Bermúdez, José. 2009. *Decision theory and rationality*. Oxford University Press.

Birnbaum, Michael H. 2008. "New paradoxes of risky decision making." *Psychological Review* 115:463–501.

Bonner, Carissa and Newell, Ben. 2010. "In conflict with ourselves? An investigation of heuristic and analytic processes in decision making." *Memory and Cognition* 38:186–196.

Bradley, Ben. 2006. "Against satisficing consequentialism." *Utilitas* 18:97–108.

Byron, Michael. 1998. "Satisficing and optimality." *Ethics* 109:67–93.

Caplin, Mark Dean and Martin, Daniel. 2011. "Search and satisficing." *American Economic Review* 101:2899–2922.

Cherniak, Christopher. 1981. "Minimal rationality." *Mind* 90:161–183.

—. 1986. *Minimal rationality*. MIT Press.

Chickering, David Maxwell. 1996. "Learning Bayesian networks is NP-complete." In Doug Fisher and Hans-J. Lenz (eds.), *Learning from data*, 121–130. Springer.

Dagum, Paul and Luby, Michael. 1993. "Approximating probabilistic inference in Bayesian belief networks is NP-hard." *Artificial Intelligence* 60:141–153.

De Neys, Wim and Glumicic, Tamara. 2008. "Conflict monitoring in dual process theories of thinking." *Cognition* 106:1248–1299.

De Neys, Wim, Rossi, Sandrine, and Houdé, Olivier. 2013. "Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools." *Psychonomic Bulletin and Review* 20:269–273.

Desvousges, William, Johnson, Reed, Dunford, Richard, Boyle, Kevin, Hudson, Sara, and Wilson, Nicole. 1992. *Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy*. Research Triangle Institute.

Elster, Jon. 1983. *Sour grapes: Studies in the subversion of rationality*. Cambridge University Press.

Frederick, Shane. 2005. "Cognitive reflection and decision making." *Journal of Economic Perspectives* 19:25–42.

Fredrickson, Barbara L and Kahneman, Daniel. 1993. "Duration neglect in retrospective evaluations of affective episodes." *Journal of Personality and Social Psychology* 65:45–55.

Friedman, Jane. forthcoming a. "The epistemic and the zetetic." *Philosophical Review* forthcoming.

Gangemi, Amelia, Bourgeois-Gironde, Sacha, and Mancini, Francesco. 2015. "Feelings of error in reasoning - in search of a phenomenon." *Thinking and Reasoning* 21:383–393.

Geman, Stuart, Bienenstock, Elie, and Doursat, René. 1992. "Neural networks and the bias/variance dilemma." *Neural Computation* 4:1–58.

Gigerenzer, Gerd and Brighton, Henry. 2009. "Homo heuristicus: Why biased minds make better inferences." *Topics in Cognitive Science* 1:107–143.

Gigerenzer, Gerd and Gaissmaier, Wolfgang. 2011. "Heuristic decision making." *Annual Review of Psychology* 62:451–482.

Gigerenzer, Gerd and Selten, Reinhard (eds.). 2001. *Bounded rationality: The adaptive toolbox*. MIT press.

Gilovich, Thomas and Griffin, Dale. 2002. "Heuristics and biases: Then and now." In Thomas Gilovich, Dale Griffin, and Daniel Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment*, 1–18. Cambridge University Press.

Hampton, Robert R. 2009. "Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms?" *Comparative Cognition and Behavior Reviews* 4:17–28.

Harsanyi, John. 1977. "Rule utilitarianism and decision theory." *Erkenntnis* 11:25–53.

Icard, Thomas. 2018. "Bayes, bounds, and rational analysis." *Philosophy of Science* 85:79–101.

Iyengar, Sheena and Lepper, Mark. 2000. "When choice is demotivating: Can one desire too much of a good thing?" *Journal of Personality and Social Psychology* 79:995–1006.

Jackson, Simon, Kleitman, Sabina, Howie, Pauline, and Stankov, Lazar. 2016. "Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases." *Frontiers in Psychology* 7:e1559.

Johnson, Eric and Payne, John. 1985. "Effort and accuracy in choice." *Management Science* 31:395–414.

Johnson, Eric, Tubau, Elisabet, and De Neys, Wim. 2016. "The doubting system 1: Evidence for automatic substitution sensitivity,." *Acta Psychologica* 164:56–64.

Kagan, Shelly. 2000. "Evaluative focal points." In Brad Hooker, Elinor Mason, and Dale E. Miller (eds.), *Morality, rules and consequences*, 134–55. Edinburgh University Press.

Kahneman, Daniel and Frederick, Shane. 2002. "Representativeness revisited: Attribute substitution in intuitive judgment." In Thomas Gilovich, Dale Griffin, and Daniel Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment*, 48–81. Cambridge University Press.

Klein, Gary. 2001. "The fiction of optimization." In Gerd Gigerenzer and Reinhard Selten (eds.), *Bounded rationality: The adaptive toolbox*, 103–121. MIT Press.

Knight, Frank. 1921. *Risk, uncertainty and profit*. University of Chicago Press.

Levi, Isaac. 1991. *The fixation of belief and its undoing: Changing beliefs through inquiry*. Cambridge University Press.

—. 2012. *Pragmatism and inquiry*. Oxford University Press.

Lieder, Falk and Griffiths, Thomas. 2017. "Strategy selection as rational metareasoning." *Psychological Review* 124:762–794.

—. forthcoming. "Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources." *Behavioral and Brain Sciences* 43.

Lin, Hanti. 2014. "On the regress problem of deciding how to decide." *Synthese* 191:661–670.

Lipman, Barton. 1991. "How to decide how to decide how to ...: Modeling limited rationality." *Econometrica* 59:1105–1125.

Lord, Errol. 2017. "What you're rationally required to do and what you ought to do (are the same thing!)." *Mind* 126:1109–1154.

Mani, Anandi, Mullainathan, Sendhil, Shafir, Eldar, and Zhao, Jiaying. 2013. "Poverty impedes cognitive function." *Science* 341:976–980.

Marewski, Julian and Schooler, Lael. 2011. "Cognitive niches: An ecological model of strategy selection." *Psychological Review* 118:393–437.

Mongin, Philippe. 2000. "Does optimization imply rationality?" *Synthese* 124:73–111.

Mongin, Philippe and Walliser, Bernard. 1988. "Infinite regressions in the optimizing theory of decision." In Bertrand Munier (ed.), *Risk, decision and rationality*, 435–457. Springer.

Mulgan, Tim. 2001. "How satisficers get away with murder." *International Journal of Philosophical Studies* 9:41–46.

Parfit, Derek. 1984. *Reasons and persons*. Oxford University Press.

Pettit, Philip. 1984. "Satisficing consequentialism." *Proceedings of the Aristotelian Society, Supplement* 58:165–176.

Pettit, Philip and Smith, Michael. 2000. "Global consequentialism." In Brad Hooker, Elinor Mason, and Dale E. Miller (eds.), *Morality, rules and consequences*, 121–133. Edinburgh University Press.

Proust, Joëlle. 2013. *The philosophy of metacognition*. Oxford University Press.

Railton, Peter. 1984. "Alienation, consequentialism, and the demands of morality." *Philosophy and Public Affairs* 13:134–171.

Russell, Stewart and Wefald, Eric. 1991. "Principles of metareasoning." *Artificial Intelligence* 49:361–395.

Russell, Stewart J., Subramanian, Devika, and Parr, Ronald. 1995. "Provably bounded optimal agents." *Journal of Artificial Intelligence Research* 2:575–609.

Scheibehenne, Benjamin, Greifeneder, Rainer, and Todd, Peter. 2010. "Can there ever be too many options? A meta-analytic review of choice overload." *Journal of Consumer Research* 37:409–425.

Schwartz, Barry, Ward, Andrew, Monterosso, John, Lyubomirsky, Sonja, White, Katherine, and Lehman, Darrin R. 2002. "Maximizing versus satisficing: Happiness is a matter of choice." *Journal of Personality and Social Psychology* 83:1178–1197.

Selten, Reinhard. 1998. "Aspiration adaptation theory." *Journal of Mathematical Psychology* 42:191–214.

Shah, Anuj, Mullainathan, Sendhil, and Shafir, Eldar. 2012. "Some consequences of having too little." *Science* 338:682–685.

Simon, Grégory, Lubin, Amélie, Houdé, Olivier, and De Neys, Wim. 2015. "Anterior cingulate cortex and intuitive bias detection during number conservation." *Cognitive Neuroscience* 6:158–168.

Simon, Herbert. 1955. "A behavioral model of rational choice." *Quarterly Journal of Economics* 69:99–118.

—. 1976. "From substantive to procedural rationality." In T.J. Kastelein, S.K. Kulpers, W.A. Nijenhuls, and R.G. Wagenaar (eds.), *25 years of economic theory*, 65–86. Springer.

Slote, Michael. 1984. "Satisficing consequentialism." *Proceedings of the Aristotelian Society, Supplement* 58:139–163.

—. 2004. "Two views of satisficing." In Michael Byron (ed.), *Satisficing and maximizing: Moral theorists on practical reason*, 14–29. Cambridge University Press.

Smith, Holly. 1991. "Deciding how to decide: Is there a regress problem?" In Michael Bachrach and S.L. Hurley (eds.), *Foundations of decision theory*, 194–217. Blackwell.

Swanton, Christine. 1993. "Satisficing and virtue." *Journal of Philosophy* 90:33–48.

—. 2004. "Satisficing and perfectionism in virtue ethics." In Michael Byron (ed.), *Satisficing and maximizing: Moral theorists on practical reason*, 176–189. Cambridge University Press.

Thompson, Valerie and Johnson, Stephen. 2014. "Conflict, metacognition, and analytic thinking." *Thinking and Reasoning* 20:215–244.

Thompson, Valerie, Turner, Jamie, and Pennycook, Gordon. 2011. "Intuition, reason, and metacognition." *Cognitive Psychology* 63:107–140.

Travers, Eoin, Rolison, Jonathan, and Feeney, Aidan. 2016. "The time course of conflict on the Cognitive Reflection Test." *Cognition* 150:109–118.

Tversky, Amos and Kahneman, Daniel. 1974. "Judgment under uncertainty: Heuristics and biases." *Science* 185:1124–1131.

Vassilakis, Spyros. 1992. "Some economic applications of Scott domains." *Mathematical Social Science* 24:173–208.

Voigt, Kristin. 2007. "The harshness objection: Is luck egalitarianism too harsh on the victims of option luck?" *Ethical Theory and Moral Practice* 10:389–407.