The Predicament of Choice  
Ralph Wedgwood

0. Rationality for finite thinkers

How is it rational for agents like us to make choices or decisions about what to do? According to classical versions of decision theory, the rational agent’s choices must maximize some kind of expectation of some kind of value. In this paper, I shall assume that some approach that is broadly in line with these classical theories is correct.¹

In several crucial ways, however, these classical theories are fantastically idealized. In particular, most of these classical theories assume that the relevant agents have an astonishingly rich system of mental states. For example, most of these theories, like the seminal theory of Leonard Savage (1954), assume that the relevant agents have complete set of preferences, defined over an infinite domain of prospects (where by “prospects”, I simply mean whatever the theory in question takes the objects of preferences to be). According to this assumption:

There is an infinite domain of prospects, such that for every pair of prospects A and B in this infinite domain, the agent either prefers A over B, or prefers B over A, or is positively indifferent between A and B.

Similarly, most of these classical theories assume that the relevant agents have a complete set of credences, defined over an infinite algebra of propositions (where an “algebra” is a set of propositions that is closed under Boolean operations like negation, and disjunction, and the like):

There is also an infinite algebra of propositions, such that the agent has a unique precise credence in every single proposition in this algebra.

Suppose that – as in theories like those of Richard C. Jeffrey (1984) and James M. Joyce (1999) – the objects of preference are themselves simply propositions, where these propositions include acts (that is, propositions to the effect that the agent performs an action of a certain sort at a certain time). Suppose also that (as seems to be implicitly assumed by these theorists) these propositions include all the propositions that can be built up out of concepts that the agent possesses. Given these suppositions, every single proposition that is built up out of concepts that the agent possesses has a definite expected value – where this expectation is defined in terms of the agent’s credences. In that sense, the agent has a view about the value of every act that can be picked out by any of the concepts that she possesses. Presumably, such an agent will choose an act only if the agent believes the act to be available, and there is no alternative act, which the agent also believes to be available, that has higher expected value.

My concern in this discussion is with agents who cannot possibly have a complete infinite system of credences or preferences of this sort. That is, I am concerned with agents whose

¹ I shall not assume that the kind of value that rational choices have to maximize the expectation of is what decision theorists call “utility”; for some arguments against this assumption, see Wedgwood (forthcoming).
system of beliefs, preferences, and other attitudes is unavoidably *gappy* – in the sense that there are some propositions built up out of concepts that the agent possesses, towards which the agent in fact has no attitudes towards at all.

Such agents can *fill in* some of these gaps, by acquiring attitudes towards some of the propositions that they previously had no attitudes towards. But filling in these gaps is always at least to some degree *costly* (at least in terms of time and mental effort, and so on). For such agents, there seem to be two possible defects in their decision making:

- **a.** The agent might make the decision too quickly, without having filled in enough of these gaps – that is, without thinking about or forming attitudes towards enough propositions.
- **b.** The agent might dither without making a decision, wasting precious time by filling in more of these gaps than is necessary, instead of making a decision in a timely way.

Under some circumstances, it seems, each of these defects can be irrational. But what is the principle of rationality that explains what rationality requires in such cases? To what extent does rationality require us to bear these costs, and to fill in these gaps, before making a decision about what to do?

There has been considerable discussion of “bounded rationality” among cognitive psychologists. Many of these cognitive psychologists focus on discovering the “heuristics” that are actually employed by the information-processing mechanisms of the human mind, which enable us to approximate to thinking rationally in spite of our limitations. My concern in this discussion is different. I shall not aim to understand how the human mind actually works. My concern is normative: How is it *rational* for gappy agents of this kind to make decisions about what to do? I shall assume that merely being a gappy agent is not in itself irrational in any way. So I shall aim to develop a kind of ideal theory: What would it be for an agent who is *limited* in this way to be *perfectly rational*?

This question has not been extensively explored by philosophers. Among the few philosophers who have explored the question, some – like Michael Slote (1989) and Jon Elster (1983: 16–18) – have proposed a *satisficing* conception of rationality, according to which it is rational to choose an option if and only if the option’s expected value is *good enough*. Intuitively, however, even if an option’s expected value is “good enough”, it does not seem rational to choose the option if one has consciously considered an alternative option that is *clearly better* than the option in question. So, I shall assume that the correct theory of rational choice will retain the classical ideal of optimization in some form.

---

2 For a general account of *inference* as the way in which such gappy agents fill in the gaps in their belief-system, see Wedgwood (2012).

3 I raised but did not attempt to solve this problem in some of my earlier work (Wedgwood 2011, 290). In this paper, I return to the overdue task of tackling this problem.

4 For a distinguished example of such work in cognitive psychology see Gigerenzer (2002, chaps. 7 and 8).
A similar point applies to Michael Bratman’s (1989) conception of rational planning. Bratman agrees with the ideal of optimization to the extent that he thinks that the initial adoption of a plan is rational only if the plan in question has optimal expected value. However, once the plan has been adopted, Bratman (1989, 72) thinks that in some circumstances it can be rational to persist with the plan even if one has thought of an alternative plan that one’s evidence reveals to be better in every way. But this seems to be a familiar form of irrationality – sticking rigidly to a preconceived plan even though one’s evidence reveals that it would have been better to be more flexible. For these reasons, the account of rational decision making that I shall develop here will be closer to the classical ideal of optimization than Bratman’s theory.

1. Thinking of options

In general, it seems, you cannot decide to take a course of action unless you have thought of – and in that sense have attitudes towards – that course of action. Similarly, you cannot decide against a course of action unless you have thought of, and so have attitudes towards, that course of action as well. In fact, in every situation in which you might act, there are infinitely many courses of action that you might take; but if you are a gappy agent of the kind that concerns us here, you will only ever think of (or have attitudes towards) a small finite subset of that vast infinite set.

In general, you need to think of each of the courses of action that you decide between as an option: you make your choice or decision precisely by comparing these options with each other, and deciding on one of the options that does not in the relevant way seem worse than any alternative. So, a crucial presupposition of every decision is the set of options that the agent is deciding between.

If the agent is a gappy agent, the set of options that she is deciding between will inevitably be finite, a tiny subset of all the infinitely many available courses of action. For this reason, it will make a big difference to the upshot of agent’s decision making what this set of options is. But where does this set of options come from?

To give a good theory of rational choice, it is essential to provide some account of the set of options that the rational agent is deciding between. Consider the theory of rational choice that says that, whatever set of options the agent happens to be deciding between, it is rational for the agent to choose a member of this set of options that has maximal expected value. This is a highly implausible theory. Suppose that the set of options that the agent is deciding between contains only single member – that is, there is precisely one course of action that the agent is treating as an option. Then it is utterly trivial that there is no alternative member of this set that has higher expected value than this option – and in this sense, it is trivial that this option has maximal expected value. So this theory would imply that whenever an agent considers only one course of action as an option, it is rational for the agent to choose the option!

---

5 The fact that the relevant set of options is finite guarantees that there is always at least one optimal element in the set. (We need not worry about sets in which for every member there is another distinct member that is even better.)
It would also not be enough just to require that the agent must consider a set of options that contains more than one member. One problem is that the agent might decide between an option-set that contains only two members, even though it is obvious that if the agent just thought for another minute before making her choice, she would think of some much better options than the two on her initial list. Intuitively, it would be irrational for the agent to make her choice between these two bad options in this way. A second problem concerns cases in which the agent decides between a huge list of options; in some of these cases, the agent might already have been irrational to have wasted as much time as she did identifying all of those many different options. So we still do not have a plausible answer to the question of what determines which set of options the rational agent will decide between.

It seems plausible that a rational agent will only include a course of action on the list of options that she is deciding between if she believes that the course of action in question is possible or available in the relevant way. One might suggest, then, that the options that the rational agent decides between are precisely those that the agent rationally believes to be available.

This solution may well be the best solution within the context of the fantastically idealized theories of rational choice that I mentioned at the outset. But it does not seem plausible if we assume that the agent’s system of beliefs and other attitudes is gappy (in the sense that I explained above). There are at least two reasons for this.

First, since the agent’s belief-system is gappy, there may only be a few courses of action that the agent believes to be available; and it may seem overwhelmingly probable that there are several much more promising courses of action that, with a bit of thought, the agent would come to believe to be available. In such cases, it might be irrational for the agent to make her choice only among the options that she currently believes to be available.

Secondly, for a course of action to be included on the list of options that one is deciding between, it is arguably not enough simply for one to have a standing background belief that the course of action is available. This belief needs to come to mind in some way; it is only if this belief comes to mind in this way that the course of action in question is considered as an option that one might decide on. Even if one already believes that a certain course of action is available, there may be further costs in bringing this belief to mind in this way. In some cases, however, there may be too many courses of action that the agent believes to be available. Some of these courses of action that the agent believes to be available might be obviously terrible, or just trivial variants of other available courses of action. In such cases, it might be quite pointless for the agent to pay the further costs of bringing these terrible or redundant courses of action to mind if she already has a big enough set of sufficiently promising options to decide between. So it seems that the set of options that the rational agent decides between need not include all of the courses of action that she believes to be available.

To keep things simple, I shall assume here that once the agent has identified a set of options that it is rational for her to decide between, the agent’s evidence, together all the other facts about

---

6 For an account of what it is for an agent to believe an option to be available, see Wedgwood (2011, 282f.).
what is going on in her mind, will determine an expected value of the relevant sort for each of these options. Then, I shall assume, the rational agent will choose between these options, by choosing an option that has maximal expected value. So, in effect, the central question for our present purposes is: How is the agent to identify a set of options that it is rational for her to decide between?

Presumably, the rational agent must go through a process of some kind that will result in a certain set of options, each of which the agent rationally believes to be available, coming to mind in the appropriate way. But what sort of process is this?

One might think of cases in which the agent carries out empirical investigations in order to identify options to decide between. For example, the agent might look through job advertisements, or listings of houses for sale, or course catalogues, or the like. In general, when the agent carries out such investigations, she deliberately acts in ways that will cause experiences from which she might learn new facts about the world – in a way that will normally result in her coming to think of options that she might decide between.

Such empirical investigations are clearly of great importance in everyday life. But for our present purposes, I shall set these cases of empirical investigations to one side. Such empirical investigations are themselves deliberate intentional actions. If one deliberately carries out such empirical investigations, one must presumably have decided to do so; and so the option of carrying out such an investigative course of action must have occurred to one. That is, one must already have thought of such an investigative course of action as one of the options to decide between. So the more fundamental problem concerns cases in which the process that results in one’s thinking of the options that one will decide between does not consist in one’s carrying out such an investigative action – or in general in one’s acting intentionally or deliberately at all.

How could it be that the process resulting in one’s thinking of the options that one will decide between does not consist in one’s acting intentionally or deliberately at all? It seems that if this process does not consist of one’s acting intentionally at all, it must be something like a process of reasoning or inference – that is, a process by means of which one teases out some of the implications that are already implicit in one’s system of beliefs and other attitudes and mental states.

For example, the agent might ask herself, “What can I do?” or “What options are available?” When the agent asks herself such questions, some possible hypotheses about what options are available might occur to her. For example, a hypothesis of the form ‘I could φ’ might occur to her. It may be that she already believes that this hypothesis is true; or if she does not yet believe it, she could engage in ordinary inferential reasoning, and thereby infer that this hypothesis is true from other propositions that she already believes. This might result in the action of her φ-ing occurring to her as one of the options that she might decide between.

In this way, this process of reasoning is initiated by an event in which a possible hypothesis about what options are available occurs to the agent. This event is an input to the agent’s process of reasoning, or a trigger that initiates this process of reasoning; but it seems that such events do not themselves directly result from any process of reasoning – at least, not from any process of
reasoning at the personal level (they may result from the way in which various sub-personal mechanisms realized in various brain processes compute various algorithms, or the like). In fact, they seem to result from an unconscious process, somewhat like the processes that are often involved when one searches one’s memory in trying to remember something.

Such processes may differ from case to case in various ways. In some cases, there are some very basic and ordinary mental dispositions that explain which options occur to one. For example, certain familiar sensations will immediately lead one to think of visiting the lavatory. In other cases, however, the explanation of which options occur to one involves the full mystery of creativity. A good example of this is supplied by the options that a great poet considers of which words to use in the next line of her poem. However, something that these processes seem to have in common is that they are unconscious processes.

Since the process that results in a hypothesis like ‘I could φ’ occurring to one is an unconscious process, and not a personal-level process of reasoning, it seems that this process does not respond to any personal-level mental states that make it rational for one to respond in this way. In this way, it seems to be a non-rational process – a process that cannot be guided by requirements of rationality, and cannot count as either rational or irrational (although such processes can undoubtedly count as fortunate or unfortunate).

Nonetheless, this process takes time, and is at least to that extent costly. So one faces a problem: How long should one allow this process of thinking of options to unfold, before one makes one’s choice? In making choices or decisions about what to do, one could delay making one’s choice to allow this process to unfold, or one could make a snap decision without allowing the process to unfold any further.

Even though the event of an option’s occurring to one results from an unconscious process, it is itself a conscious mental event. One can respond to such conscious events by either making a choice between the options that have already occurred to one, or not making a choice yet, and waiting for more such options to occur to one. When one responds to such events in either of these ways, one is in effect reasoning. This kind of reasoning can be done in either a rational or an irrational manner. Our task is to understand what this difference amounts to.

2. Deciding how to decide?

Sometimes, we might actually decide how to decide; we might make a decision about how much time to devote to thinking of options before making the decision between those options. There is no doubt that we can and do sometimes make such higher-order decisions about how to decide; and sometimes it seems quite rational for us to make such higher-order decisions.

However, it would absurd to suggest that we are always rationally required first to decide how to decide before we ever make any decision. Complying with this suggested requirement would involve an infinite regress – before making any decision, one would have to have made a distinct prior decision, and so on ad infinitum. So this suggestion makes rational decisions impossible.
It would be less absurd to suggest that before every “substantive” or “first-order” decision (that is, every decision between courses of action that do not themselves consist purely in deliberative mental actions), one must first make a “procedural” or “higher-order” decision about how to make that “substantive” decision. This second suggested requirement does not give rise to an infinite regress in the same way; it appears possible, at least in principle, to comply with this requirement.\(^7\)

However, even though it would be possible to comply with this suggested requirement, it seems not to solve our problem. The basic problem is simple. The “procedural” decision also presupposes thinking of a set of options to decide between, just as much as the “substantive” decision. So the very same problem arises at the level of the higher-order or procedural decision as at the level of the first-order or substantive decision.

Thus, you might make this “procedural” decision in an irrational way – either by not identifying enough options, or by wasting too much time thinking of options after you have already thought of enough. For example, you might irrationally decide only between the options *Making your decision right now, among the options that you have already thought of* and *Spending another 24 hours trying to think of further options to decide between*, without considering other options like *Spending another 5 minutes trying to think of further options to decide between*. So this suggestion cannot provide a sufficiently general answer to the question of how much time rationality requires us to spend on thinking of options to decide between.

In this way, then, this suggested requirement does not solve our problem. Since there seems to be no other reason for thinking that rationality necessarily involves satisfying this requirement, it seems doubtful whether there is any such requirement at all. It seems intuitively plausible that there are cases in which it is perfectly rational to make the substantive “first-order” decision without first making any procedural “higher-order” decision.

In general, the idea that we should solve our problem by appealing to higher-order decisions about how to make one’s first-order decision seems to be an excessively over-intellectualized approach to the problem.\(^8\) Even if the correct solution does involve some sensitivity to such higher-order considerations, we need to find a different interpretation of this sensitivity than simply as yet another decision that needs to be made before making the first-order or substantive decision.

3. **Is Thinking of more options itself just another option?**

It might be thought that the problem with the previous suggestion was that it separated the higher-order decision from the first-order decision, treating them as two decisions. So, an alternative suggestion would be that the first-order options and the higher-order options should

---

\(^7\) No philosopher, to my knowledge, has proposed anything like this suggested requirement. For an illuminating discussion of such higher-order processes of deciding how to decide, see Lin (2014).

\(^8\) For a compelling argument against such over-intellectualized conceptions of rational agency, see Arpaly (2003).
be combined into a single decision. That is, in effect, according to this suggestion, *thinking of more options* is itself just another option, and that rational agents will always include this option – the option of thinking of more options – as one of the options that they will decide between. In other words, according to this suggestion, the rational agent’s set of options always include the option of *not deciding yet but going “back to the drawing-board”*. This suggestion could form the basis of the following account of rational choice: according to this account, it is rational to choose an option, out of a set that involves this “back to the drawing-board” option, just in case the choice maximizes the relevant sort of expectation of the relevant sort of value, when compared with other members of this set.

However, according to this suggested account, the options that the agent is deciding between will not typically be *alternatives* to each other – that is, it is not rational for the agent to be certain that she will not take *more than one* of these options. For example, suppose that the agent’s options are: $A$, $B$, and *Trying to think of more options*. Even if the agent takes the third option, and tries to think of more options, she might end up taking option $A$ after all.

But it does seem that within the kind of approach to rational choice that is centred around the idea of maximizing expected value, it is crucial that the options that a rational agent is deciding between should be alternatives to each other in this way. According to this central idea, it is rational to choose an option only if *no alternative* has greater expected value. But if the options are not alternatives, we cannot infer from the fact that one option $A$ has lower expected value than another option $B$ that it is not rational to choose option $A$.

This point arises for the following reason. On many conceptions of “expected value”, it can happen that there are two options such that one option has higher expected value than the other, even if intuitively the rational agent will in fact choose *both* of these options. For example, it could happen that an option $A \& B$ has greater expected value than $A$. (This could be the case, for example, if (i) there is a significant chance that if one chooses $A$ one will also choose $\neg B$, and (ii) $A \& \neg B$ has a lower expected value than $A \& B$.) But presumably, it does not follow that because $A$ has lower expected value than another option $A \& B$, it is irrational for the agent to choose $A$ for this reason. On the contrary, if it is rational for the agent to choose $A \& B$, it is presumably also rational to choose $A$. For this reason, within the framework of the kind of approach to rational choice that we are exploring here, we need the options that the agent is deciding between to be alternatives to each other.

Thus, this approach cannot simply combine first-order and higher-order options in the list of options that the agent is deciding between; instead, to make this approach to work, every single option must combine both first-order and higher-order elements. In effect, this approach would have to say that the rational agent must always decide between a set of options of the form: $A$

---

9 The problem is not restricted to cases where one of these two distinct prospects actually entails the other (as $A \& B$ entails $A$). $A \lor B$ might have higher expected value than $B \lor C$, and yet it might be rational to choose both.
without trying to think of more options, B without trying to thinking of more options, ... Trying to think of more options.

But this approach to rational choice seems to be too restrictive. It implies that it cannot be rational for an agent to make any choice unless the agent conceives of the options that she is deciding between in this extremely specific way. In effect, it requires the thinker to think of every option partly in terms of the role of her own thinking. This seems to demand an excessively reflective way of making decisions: on this view, a rational agent, throughout all her decision making, has to deploy psychological concepts and think about her own thinking. But surely a rational agent’s attention can be fixed on what she might do in the external world; she does not have to think all the time about her own thinking as well. It seems more plausible that the factors that ensure that the agent is spending a rationally appropriate amount of time on trying to think of options may often be in the background of her decision making itself, rather than in the foreground, as part of the options that the agent is deciding between.¹⁰

As I have already suggested, the rational agent needs in a way to be sensitive to these higher-order considerations. But we need to find a way of understanding how this sensitivity need not take the form of practically deliberating and deciding between a set of higher-order options at all. This is the kind of solution that I shall try to outline in the following section.

4. The proposed solution

At every given time, you have a certain body of evidence, and a certain set of mental states. As I shall explain, it may be that this evidence and these mental states somehow determine a set of probability functions – the probability functions that rationally should guide your thinking at that time. If there is such a set of probability functions, then, as I shall also explain, this will help us to solve the problem that we are concerned with.

The agent’s actual belief-system, I am assuming, is gappy: there are propositions towards which the agent is quite capable of having attitudes, but in fact has no attitudes at all. A probability function, on the other hand, cannot be gappy in this way. A probability function is defined over a whole field of propositions, which must take the form of a propositional algebra –that is, it must be closed under Boolean operations like negation and disjunction.

Moreover, we should presumably also allow that the agent’s actual belief-system may be not only gappy, but probabilistically incoherent as well. Thus, even if the agent’s evidence and all the agent’s other mental states do determine a set of probability functions, as the probability functions that rationally should guide the agent’s thinking at the relevant time, this need not be a set of probability functions that simply corresponds to the belief-system that the agent happens to have.

¹⁰ For an illuminating argument for the importance of understanding that the factors that make processes of decision making rational may be “backgrounded” in this way, see Pettit and Smith (1990).
So how could the agent’s evidence and set of mental states determine a set of probability functions? Several different accounts could be offered at this point, but to fix ideas, I propose the following. There may be many synchronic requirements that an agent’s set of mental states needs to meet to be fully rational. Since these are all requirements about how the mental states that the agent has at a particular time should fit together, these requirements are all broadly speaking requirements of coherence. These may include requirements, not just about how one’s beliefs and credences should cohere with each other, but also about how one’s beliefs should cohere with one’s experiences and episodic memories and the like.\(^\text{11}\)

We can distinguish between some mental states (like beliefs) that can be revised through reasoning, and other states (like experiences and episodic memories) that cannot be revised through reasoning in this way. Then consider the set of probability functions nearest to the agent’s actual belief-system each of which corresponds to a credence function that meets all requirements of coherence both internally (between the credences that the function assigns to all the propositions for which it is defined) and with the agent’s mental states (like her experiences and episodic memories) that cannot be revised through reasoning. To fix ideas, let us suppose that it is this set of probability functions that should rationally guide the agent at the time in question.

Now, at every time \(t\), there are two particularly crucial higher-order options: Deciding then (at \(t\)) between the first-order options that one has already thought of, and Not deciding then (at \(t\)) but trying to think of more options.\(^\text{12}\) Presumably, these two higher-order options can be presented in thought by means of propositions. In particular, at \(t\) itself these options can be presented in thought by means distinctively indexical first-personal propositions: in these propositions, the time \(t\) is presented indexically as now, and the agent is presented first-personally as me. I shall assume that even if the agent herself has no attitudes at all these propositions, these propositions form part of the field of propositions over which the probability functions determined by the agent’s evidence and other mental states are defined.

Another kind of proposition that is crucial to rational decision-making are propositions that specify how good or how desirable the agent’s behaviour is during a certain period of time \(T\). (For present purposes, we need not worry about exactly what sense of ‘good’ or ‘desirable’ is in question here. To fix ideas, suppose that the degree to which an agent’s behaviour is good during a given period \(T\) is fixed by how much reason there is all things considered in favour of that behaviour.) Each of these propositions specifies the degree to which the agent’s behaviour during \(T\) is good or desirable. I shall assume here that these degrees of goodness can be measured on

\(^{11}\) For a view that is effectively equivalent to postulating such requirements of coherence between beliefs and experiences, see Haack (1994).

\(^{12}\) Obviously, these two options are not strictly exhaustive. There are further alternatives as well: for example, one might take neither option, neither deciding at \(t\) nor thinking of any more options (but for example simply postponing the decision to a later occasion). To keep things simple, however, I shall assume that the two options mentioned in the main text above are in fact exhaustive. A full account of the topic would have to dispense with this assumption, but I shall ignore these complications here.
some kind of scale. The unit and perhaps also the zero point on this scale may be completely arbitrary, but suppose that our theory of rational decision-making has picked on one of these scales, to represent the content of these propositions. So, these propositions can be represented by formulae of the form ‘\( V(T) = n \)’, where \( n \) is the measure on this scale of how good or desirable the agent’s behaviour is during that time \( T \).

At any time \( t \) at which the agent deliberates about what to do, a certain stretch of the agent’s future behaviour is potentially under the control of the thinking that the agent engages in at that time; let us refer to the period of time occupied by this stretch of behaviour as ‘\( T_t \)’.

Suppose that there is a partition of propositions of the form ‘\( V(T_t) = n \)’, in the sense of a set of such propositions such that the relevant probability function is certain that exactly one member of this set of propositions is true (that is, the relevant probability function assigns probability 1 to the disjunction of all these propositions, and probability 0 to the conjunction of any two of these propositions). For every option \( A \), let \( A \) be the relevant (first-personal indexical) proposition that the agent takes that option \( A \) at \( t \). Then the relevant expected value of an option \( A \) according to a probability \( P(\bullet) \) can be defined as:

\[
\sum_a n \times P(V(T_t) = n|A).
\]

Now, suppose that according to every probability function in the set that is determined by the agent’s evidence and other mental states, one of these two options – either Deciding then (at \( t \)) between the first-order options that one has already thought of, or Not deciding then (at \( t \)) but trying to think of more options – has higher expected value in this sense than the other. Then, as I shall put, this option has definitely higher expected value than the other.

In this way, the agent’s evidence and other mental states determine what this set of probability functions is, and this set of probability functions determines whether or not one of these options has definitely higher expected value than the other. So, by transitivity, the agent’s evidence and other mental states determine whether or not it is the case that that option has definitely higher expected value than the other.

It seems that it is possible for you, if you are rational, to respond directly to the fact about your evidence and other mental states that determines that one of these two higher-order options has definitely lower expected value than the other by not taking the former higher-order option. Similarly, if you are rational, you can respond to the fact that determines that one of these higher-order options has definitely higher expected value than every alternative by taking that higher-order option.

In saying that you can respond “directly” to the fact that this option has definitely higher expected value, I mean that you do not have to make any kind of choice or decision against taking the relevant higher-order option in order to respond in this way; at least to the extent that you are fully rational, you can respond to such facts about these higher-order options by directly taking one of these options, and not taking the other option, without any further conscious reasoning or deliberation.
Indeed, given the way in which I have characterized what it is for one of these options to have definitely lower expected value than the other, we can say, more strongly, that you do not even need to have any thoughts at all about these higher-order options. It is enough if you respond in the appropriate way to the evidence and other mental states that determine that this option has definitely lower expected value. You can take these options without thinking about them at all. This is because to “take” the option of Deciding now between the first-order options that you have already thought of is simply to make a decision between the relevant options at the relevant time, and to “take” the option of Not deciding now and trying to think of more options is simply not to make a decision at that time, and to try to think of more options instead.

You can also “respond” to the fact that one of these options has a lower expected value, not by thinking explicitly about probabilities, but simply by responding directly to the facts about your evidence and other mental states that determine the expected values of these options. In responding to the evidence, you also do not have to think about the evidence; it is required only that you have this evidence, and that you respond to the evidence in a way that is explained by the fact that this evidence makes it the case that the higher-order option in question has definitely lower expected value than the alternative.

Thus, in some cases, your response to this evidence simply consists of your “taking” one of these options. In these cases, the process that leads from the possession of this evidence to your response will be direct, rather than a process of reasoning through any intervening steps. If it is a direct mental process of this sort, the agent herself will probably not have much insight into why her response is rational. According to my proposals, this does not matter. Her response can be rational even if she does not have any such insight into why it is rational.

This makes it possible to propose the following general account of rational choice: it is rational for you to decide at a time $t$ to take an option $A$ if, and only if, (i) $A$ has maximal expected value, out of all the options that you (a) have actually thought of, and (b) rationally believe to be available to you at $t$, and (ii) the higher-order option of Deciding at $t$ between the options that you have already thought of does not have definitely lower expected value than the alternative Not deciding at $t$ and trying to think of more options.

This then is my proposed solution to the problem: Rational agents will exhibit a kind of sensitivity to the expected values of these higher-order options, but that sensitivity need not take the form of any kind of explicit higher-order reasoning. As Philip Pettit and Michael Smith (1990) would put it, this sensitivity is a “backgrounded” component of rational decision making.

5. Isn’t this solution also excessively idealized?

Someone might object to the proposal that I have just made, along the following lines: This solution requires that rational agents can respond in a certain way to facts about their evidence even without thinking either about that way of responding, or about the evidence itself. Doesn’t this feature of the solution make it just as fantastically idealized as classical decision theory?

The gist of my reply to this objection is as follows. As I explained at the outset, the account that we are seeking is a kind of “ideal theory”; specifically, it is a theory of perfect rationality for
limited agents. It certainly should not be expected that such perfect rationality will be easy. On the contrary, it is a lofty ideal to which we may approximate, but which we are unlikely ever to achieve for more than a few scattered moments. I have not been seeking an account that is not an ideal theory in this sense; I have only been seeking a theory that applies to agents who are inevitably limited – in the sense that their systems of beliefs and preferences are “gappy” in the sense that I described at the outset.

In fact, all probabilistic theories of rationality must be interpreted as “ideal theories” of this sort. Consider how most probabilistic theories of rationality have attempted to account for the way in which it is rational for agents to revise their system of beliefs in response to acquiring new evidence – namely, by appeal to conditionalization. To revise your beliefs by conditionalization, you would have to shift your degree of belief in every proposition $p$ that you have any degree of belief in, to the result of conditionalizing your prior degree of belief in $p$ on whatever evidence you have just acquired. So, to revise your beliefs in this way, you would have to adjust your degrees of belief in a huge number of propositions simultaneously. If we are ever capable of doing this, we cannot do it by explicitly thinking about all those propositions at the same time.\(^\text{13}\)

We would have to be able to revise our beliefs in these propositions without thinking about what we are doing, and without thinking about the bearing of our evidence on each member of that huge set of propositions.

A similar point applies to the phenomenon of inference. Inference can be thought of as a process whereby gappy agents fill in some of the gaps in their system of beliefs. But within the context of a probabilistic theory of rationality, agents can draw inferences from premises in which they do not have maximum confidence, but only a partial degree of belief. In these cases, the degree of belief that these agents rationally should have in the conclusion of the inference is not determined purely by the degrees of belief that they have in the premises. (At most, the degrees of belief that these agents have in premises imposes a lower limit on how low their degree of belief in the conclusion can be; but rationality may in fact require a degree of belief in this conclusion that is significantly higher than this lower limit.) In these cases, the degree of belief that it is rational for the agent to have in the conclusion of the inference is determined holistically by the totality of the agent’s evidence and their whole system of beliefs.\(^\text{14}\)

In adjusting their degree of belief in this way, these agents typically will think about – and have attitudes towards – the conclusion of the inference itself; but they need not think about the way in which this conclusion relates to the totality of their current system of evidence and their whole system of beliefs.

Some philosophers defend a conception of reasoning according to which reasoning is essentially active – something that we actively do, rather than something that happens to us because of

\[^{13}\text{It is for precisely this reason that Harman (1986, 25–7) doubts that it is realistically possible for agents like us to update our beliefs by conditionalization at all.}\]

\[^{14}\text{For this point, see Wedgwood (2012, 286).}\]
automatic processes that we go through.\textsuperscript{15} It is not completely clear what is meant by this contrast between “active” and “passive” – given that it is not assumed that only intentional voluntary actions count as “active” in the relevant sense. But the paradigm case of such “active” reasoning is often supposed to consist in explicitly running through a sequence of sentences that constitutes a valid argument in one’s head – as if one were silently reciting an incantation or the like.

The process whereby an agent responds appropriately to the fact about her evidence and her other mental states that makes it the case that one of these higher-order options has definitely higher expected value than the other is in effect a form of reasoning. But it seems not to resemble these paradigm cases of “active” reasoning: on the contrary, in all likelihood, it involves all sorts of unconscious automatic processes of various kinds.\textsuperscript{16}

Occasionally, our thinking proceeds in a certain way because we have engaged in reflective thought about its proceeding in that way. Clearly, however, such cases are of necessity the exception rather than the rule. Normally, our thinking proceeds as it does without our engaging in any reflective thought about its proceeding in that way. In these normal cases, it seems that we are guided, not by reflective thought about how our thinking should proceed, but by simply certain habits or dispositions. In practical reasoning, it is these dispositions that explain how long we spend trying to think of options, and when we stop trying to think of options and make a decision between the options that we have thought of so far. According to the proposals that I am making here, the dispositions of the perfectly rational agent would non-accidentally lead the agent to conform to the principle that I have articulated at the end of the previous section.

Unfortunately, however, it should not be expected that human agents like you and me will ever be perfectly rational. At best, we may approximate to conforming to these requirements of perfect rationality. Although our practical reasoning is guided by our habits or dispositions, in the manner that I have described, our habits are unlikely to be ones that it is always rational for us to rely on. Most of us have some reason to believe that our decision-making habits are defective in some respects. For example, we know that we tend to make some decisions too quickly, without considering enough options, whereas on other questions we dither pointlessly wasting time.

To be as rational as we can, agents like us will have to compensate for these defects in our dispositions as best we can. Specifically, what agents like us need, it seems, is a sort of alarm signal that will be sensitive to the signs of our dispositions’ leading us astray; we can then rely on our dispositions except when alerted by the alarm signal, in which case we can deliberate explicitly about how to deliberate. In this way, imperfectly rational agents like us will be able at least to approximate to conforming to the requirements of rationality.

\textsuperscript{15} This conception of reasoning has been prominently defended by John Broome (2013, chaps. 12 and 13).

\textsuperscript{16} For a conception of reasoning that fits with this picture, see especially Staffel (2013).
These points, I suggest, are sufficient to answer the present objection. No one ever said that being rational was going to be easy. It is enough if it at least possible – in the sense of being compatible with our being gappy agents of the kind that I described at the outset.

6. Conclusion

Our intentional voluntary actions seem to be a paradigm example of cases in which we are active and in control of what we do; and such actions count as intentional precisely they issue from our engaging in practical reasoning. However, such conscious practical reasoning presupposes unconscious mental processes, which involve the unconscious operation of various mental dispositions.

There are two main kinds of such unconscious mental processes that I have identified here.

- The unconscious operations of the first kind of these mental dispositions – like the processes that result in options’ consciously occurring to one – are non-rational, in that they do not involve responding to requirements of rationality.
- The unconscious operations of the second kind of mental dispositions – like our dispositions to stop or to continue trying to identify more options to decide between – do respond to requirements of rationality (even though all that we are consciously aware of is making a choice after spending a certain amount of time trying to think of what options are available).

Both kinds of unconscious process are crucial for our success as rational agents – even the non-rational processes of the first kind.

Indeed, the virtue that Aristotle calls phronēsis or practical wisdom (Nicomachean Ethics 1140a24–32) may consist precisely in having good dispositions of these kinds, and especially the second kind. That is, agents who are phronimoi are especially perceptive and good at spotting opportunities for fine actions in the concrete situations that they find themselves in. This is a kind of creativity – a disposition to come up with good ideas about what one might do – and like other forms of creativity, it seems incapable of being analysed as consisting in conformity to rules or requirements of any kind.

At all events, however, whether we are phronimoi or not, the predicament of choice is that all of us, however rational we are, have to choose between the options that have occurred to us as a result of the operation of these non-rational dispositions.¹⁷

¹⁷ Earlier versions of this paper were presented at a conference on the significance of intentions at the University of Leeds in June 2014, and at the Chapel Hill Metaethics Workshop in October 2015. I am grateful to both members of both those audiences for helpful comments.
References


