

# 1

## Rationality in Retrospect

*Mark Schroeder*

It is relatively common to distinguish between two perspectives on normative appraisal of action: the objective and the subjective. Sometimes said to correspond to each of two readings of ‘ought’ claims in context, the objective perspective is that from which we are concerned with which choice is correct, and the subjective perspective is that from which we are concerned with which choice is rational. The divide between rationality and correctness comes out particularly sharply when we examine choices which, once made, do not turn out as well as hoped. After calling but losing in poker, we can see in retrospect that our choice was not, after all, the correct one in the situation—after all we lost. But nevertheless, we steel ourselves to make it again when confronted with similar evidence in prospect, because after all it was the rational choice to make.

Examples like these—and the fact that the contrast between rationality and correctness comes out most sharply in light of them—encourage a kind of picture thinking according to which the subjective perspective within which we evaluate actions as rational or irrational is the proper *prospective* perspective on choice, while the objective perspective from which we evaluate actions as correct or incorrect is the proper *retrospective* perspective on choice. Call this, for want of restraint in persuasive definition, the *naïve* picture.

It is the purpose of this chapter to demonstrate that the naïve picture gets things precisely the wrong way around. The objective perspective, I will be arguing, is the proper *prospective* perspective to take on choice, and the subjective perspective is the proper perspective of *retrospection*. In contrast to the naïve picture, we may call this the *unintuitive* picture.

## 2 RATIONALITY IN RETROSPECT

As unintuitive as my thesis sounds in prospect, I will show that this mistake has had pervasive consequences for our understanding of the relationship between rationality and correctness, for whether rationality is in any interesting way connected to praise and blame or supervenes on the internal, and for the even more basic question of whether there is any useful distinction to be drawn at all between two equally important perspectives of normative appraisal of action in the first place. My aim is to make my thesis as obvious in retrospect as it is unintuitive in prospect.

### 1.1 Distinguishing Two Dimensions of Assessment

According to the orthodox view of which I aim to give a novel defense in this chapter, what it is rational to do corresponds to what it makes sense to do given the information that we have, and we are to be blamed for doing what is irrational and praised for its contrary. Whereas what it is correct to do corresponds to what it makes sense to do in light of the facts, independently of what the agent believes about those facts, and we are to be congratulated for doing what is correct and commiserated with for doing what is not. The distinction between what is rational and what is correct arises, according to this orthodoxy, because the information that each of us has to act on can come apart from the facts.

Orthodoxy	Correctness	Rationality
In-rules	Depends on facts independently of agent's information	Depends on agent's information, independently of its truth
Out-rules	correct → merits congratulation incorrect → merits commiseration	rational → merits praise irrational → merits blame

The most obvious way in which our information can come apart from the facts is by being incomplete. A standard philosophy parable that tells this story is what has come to be known following Regan (1980) as the

miner's case or, more conveniently for my purposes (since it generalizes more easily as we will need to do later), the three envelope problem.<sup>1</sup> In this case, Amy has to choose which of three envelopes to take. She will get to keep whatever is in the envelope that she chooses. She knows that the first envelope contains \$1000, that one of the envelopes contains \$1500, and that one of them is empty. But she does not know which envelope contains the \$1500 and which is empty. So the information available to her comes apart from the totality of the facts because it is incomplete. There is a relevant fact—namely, which of the second or the third envelope contains the \$1500 (it is the second, by the way)—that she does not know. And as a result, the correct choice comes apart for her, from the rational one.<sup>2</sup>

The correct choice, in this case, is to take the second envelope. It is the one with the most money, and if we had the opportunity to advise her, this is the one that we should tell her to take. It even makes sense for Amy to say, after choosing the first envelope and then being shown that the second envelope has more money in it, 'I should have taken that one!' But it would not be rational for Amy to take the second envelope. The only reasonable choice is for her to take the first one. If she instead takes the second one, we will congratulate her but blame her for her choice. (If the first envelope had contained only \$200, we would congratulate but not blame her for this choice.) And if she takes the third, we will both blame her and commiserate with her. Congratulations are appropriate for choosing the correct thing, but praise is appropriate for choosing the rational one.

But information can also come apart from the facts by being false. One of the standard philosophical parables that tells this story is Bernard Williams' (1981) gin and tonic case. In that case, Bernie's usual bartender has given him a glass that by all appearances is the gin and tonic that he

<sup>1</sup> In the form of the three envelope problem this case comes from Ross (2006).

<sup>2</sup> I will sometimes simplify in what follows and write as if all that matters is the comparative expected value of each option, but the charitable reader will take this as proxy for whatever choice rule is the correct one for decision under uncertainty. All of the points that I make in this chapter can be made under any assumptions under which uncertainty matters for rational choice.

#### 4 RATIONALITY IN RETROSPECT

asked for, and he believes it to be so. But unluckily for Bernie, his glass does not actually contain gin and tonic, but instead is filled with gasoline. So the information available to Bernie is not entirely true. It contains the falsehood that his glass contains gin and tonic. And this is a relevant fact. As a result, the correct choice of whether to take a sip comes apart, for Bernie, from the rational one.

The correct choice is not to take a sip. If we had the opportunity to advise Bernie, this is certainly what we should tell him to do. It even makes sense for him to say, after taking a sip and discovering that it is gasoline, ‘I shouldn’t have drunk that!’ But it would not be rational for Bernie not to take a sip. After all, he has already paid for his drink and he believes it is what he asked for. If he does not take a sip, we will congratulate him, but blame him. If he does take a sip, we will praise his clear thinking but commiserate with him over its results.

This orthodox view has come under fire from multiple directions. Subjectivists and perspectivalists have argued that what you *ought* to do is really tied more closely to your perspective after all—so rationality and correctness do not come so far apart, because both are tied in some way to your information at the time of choice.<sup>3</sup> Externalists have argued that what it is rational for you to do does not supervene on your mental life, but can depend on what you *know* or are *in a position* to know—thereby rejecting, in particular, cases like the gin and tonic case altogether.<sup>4</sup> And some have even combined these two moves—arguing that rationality and what you ought to do actually *coincide* because what is rational depends *more* on the facts than on the orthodox view, while what is correct depends *less*.<sup>5</sup>

Despite these prominent dissensions from the orthodox view, however, it is, I think, more or less correct as stated. And it is an important part of the purpose of this chapter to defend it—particularly from these challenges. But the orthodox view is also commonly and easily tied up with a kind of naïve picture thinking about *why* it is correct. This naïve picture, I will endeavor to show in this chapter, is not only wrong; it is at

<sup>3</sup> Gibbons (2009, 2013); Ross (2012).

<sup>4</sup> Williamson (2000); Lord (2010).

<sup>5</sup> Kieseewetter (2016, 2017); Lord (2018).

the heart of the most important objections to the orthodox view. So getting over this naïve picture is the key to appreciating the strength of the orthodox view.

The naïve picture is simple. It says that judgments of rationality serve the purpose of *prospective* assessments of choice, whereas judgments of correctness serve the purpose of *retrospection*. The picture is that when we look forward to making a choice, it is helpful—even a guide to choice—to think about which choice is rational or most rational. Whereas it is only when looking backwards at what we have done and whether the world cooperated that it is useful to think about which choice was correct.

## 1.2 Motivating the Naïve Picture

It is worth distinguishing at least three different sources of the naïve picture. The first and simplest of these comes from simple consideration of the decisions of the poker player—a paradigmatic case of decision in the face of imperfect information. Prospectively, when I am making the decision whether to call or fold, it sure would be nice to know what cards my opponent is holding, but unfortunately I simply don't. I have to go by the information that is available to me. So when I'm deciding what to do, it is the dimension of assessment that depends on the information that is available to me—rationality—that it seems should matter to me. When I decide to call and say 'I should call', on this view, what I say is true, because I am speaking about what I ought rationally to do.

And similarly in retrospect, if I am a good enough player then I already know whether my decisions in the game were rational or not. What I learn by retrospectively on the outcome of those choices is whether they were in fact *correct*—the mode of assessment that depends on how things really were, independently of what I believed or suspected about them. So when I say, 'shoot, I should have folded' upon seeing my opponent's hand, again what I say is true, but this time because I am speaking about what I ought to do given the facts. Together, these two halves of the poker player's reasoning therefore provide elegant intuitive support to the naïve picture.

## 6 RATIONALITY IN RETROSPECT

But the intuitive support provided by the poker player's decision is strengthened considerably by a second leg of support for the naïve picture that comes from three envelope cases. In three envelope cases, not only do you not know, prospectively, which choice is the correct one; there is a very simple and difficult-to-resist argument that you are actually in a position to know that it is incorrect to take the first envelope! For from her prospective perspective, Amy can reason as follows:

- P1 Either the \$1500 is in envelope 2 or the \$1500 is in envelope 3.
- P2 If the \$1500 is in envelope 2, then choosing 2 is correct.
- P3 If choosing 2 is correct, then choosing 1 is incorrect.
- P3 If the \$1500 is in envelope 3, then choosing 3 is correct.
- P4 If choosing 3 is correct, then choosing 1 is incorrect.
- C So either way, choosing 1 is incorrect.

This argument requires only reasoning by cases, and two applications of *modus ponens* in each subproof. So as arguments go, it looks pretty awesome. It seems that Amy is, in fact, in a position to know that choosing envelope 1 is *incorrect*.

The three envelope problem therefore seems to provide a powerful reason to think that correctness not only is not *always* helpful to think about from the prospective perspective of choice, but that it can actually be a positive hindrance. In the three envelope case, Amy can know for certain that choosing envelope 1 is incorrect, but that doesn't help her *at all* to decide what to do—she certainly can't rule out taking envelope 1 on that basis. Indeed, it is completely irrelevant to her choice, since she will arrive at the same conclusion by the same reasoning no matter whether envelope 1 contains \$1000 or \$200. So the three envelope case props up the intuitive case that correctness *cannot* be the appropriate mode of assessment to consider from the prospective perspective on choice.

Now, if correctness is not a helpful guide from the prospective perspective, there may, of course, be something else that is. And the third source of the naïve picture is the idea that it is not hard to see what this might be. Since what is rational depends only on your own mental states, and not on the world independently of you, you are in a uniquely good position to be able to know what it is rational for you to do (so the third

source reasoning goes). So if correctness is so unhelpful because it is so hard to know the determinants of what it is correct for you to do, rationality is particularly helpful because it is so easy to know the determinants of what it is rational for you to do.

This third source of the naïve picture is very closely related to a family of arguments that we might accurately dub the *classical* argument for internalism about rationality. A representative version of the classical argument goes something like this:

- P1 You are always rightly blamed for making an irrational choice.
- P2 It is right to blame you for making an irrational choice only if you were in a position to know that that choice is irrational.
- P3 If what it is rational for you to do can depend on features outside of you, then you are not always in a position to know that an irrational choice is irrational.
- C What it is rational for you to do cannot depend on features outside of you.

Like the third source of motivation for the naïve picture, the classical argument for internalism about rationality assumes that you *are* in a position to know whether you satisfy the internal determinants of rationality. Otherwise the argument would lead instead to the conclusion that no choice can be irrational.

And the classical argument for internalism about rationality is also very closely related to the naïve picture itself. Premise P2 of the classical argument says that it is right to blame you for making an irrational choice only if you were in a position to know that that choice is irrational. This premise builds in the picture that to avoid blame in action, the thing to do is to focus on doing only what you know to be rational. If you are focused on this, then the goal that guides your choice at the time of action is the goal of doing the rational thing. And this, of course, is a very strong version of the idea that rationality is the proper prospective perspective on choice—that it should be the overriding goal on which you act and filter through which you evaluate your own actions in order to select among them.

## 8 RATIONALITY IN RETROSPECT

### 2.1 The Challenge to Orthodoxy

The boldest challenge to the orthodox view comes from Timothy Williamson’s anti-luminosity argument. Williamson’s anti-luminosity argument purports to show that there is no internal condition such that we are always in a position to know that we are in that condition, when we are in it. Williamson’s idea is that every condition has some boundary cases—cases that are just over the border into satisfying the condition rather than not. But Williamson thinks that knowledge satisfies a margin-for-error principle, so that knowing that something satisfies a condition requires being able to rule out nearby cases in which it does not satisfy that condition. So if you satisfy some internal condition but just barely—so that you are in a boundary case—then your position is too close to cases of not satisfying the condition to be able to rule out that that is your lot. Hence, Williamson thinks, there is no internal condition such that merely being in it guarantees that you are in a position to know that you are in it.<sup>6</sup>

Let’s take the conclusion of Williamson’s anti-luminosity argument on board. Whether or not we accept the exact reasoning in the argument, I think its conclusion is correct, and indeed that any reasonable conception of what counts in the relevant sense as someone’s information will have to allow that what your information is can come apart from what your information is about your information. And enough has been written about it elsewhere for us to explore its consequences without getting distracted by other ways of resisting.<sup>7</sup> As Williamson and others following him have noted, the anti-luminosity argument makes trouble for the classical argument for internalism about rationality.

The trouble is that if what it is rational for you to do supervenes on your wholly internal mental states, then there is some internal condition that is necessary and sufficient, in order for some choice to be irrational. But according to the anti-luminosity argument, this internal condition is

<sup>6</sup> Williamson (2000), especially pages 96–108.

<sup>7</sup> Compare Brueckner (2002); Blackson (2007); Berker (2008); and Ramachandran (2009) against, and Srinivasan (2015) for.



not luminous—you can satisfy it without being in a position to know that you do. Hence, since irrationality engenders blame, it can be right to blame you for taking that choice even though you were in no position to know that it was irrational. The argument is unsound.

Williamson and at least some of his followers have concluded that there is no good argument for internalism about rationality. What it is rational to do or to believe may depend on features of your past or your environment—even features that are unknown or unknowable to you. There is nothing particularly strange about this, they contend, because the idea that rationality could even possibly depend only on things that you are in a position to know was a Cartesian illusion.

In the hands of both Williamson and many others who have been influenced by him, this turns into the thought that what it is rational for you to do or to believe can depend, in particular, on what you know—so having information is still necessary for rationality, but the mode of possessing information that is relevant is knowledge.<sup>8</sup> Since knowledge is factive, it is a consequence of this alternative picture that rationality can only come apart from correctness by means of the incompleteness of your information—not because your information is false in any way. So this challenge has led many to think that rationality is more objective than orthodoxy allows, threatening the orthodox idea that rationality depends only on the agent's information, independently of whether it is true. And since many who accept this argument continue to accept premise P2, they are led in addition to reject P1, giving up the other orthodox commitment about rationality—that all irrational actions merit a kind of blame.

In the remainder of this chapter, I am going to provide answers to this challenge to the orthodox picture. The key to my answer lies in rejecting the naïve picture and embracing instead the prospectively unintuitive picture that rationality is the appropriate perspective for *retrospective* assessment of choice, while correctness (and 'ought') are appropriate for assessing choices from the *prospective* perspective.

<sup>8</sup> Lord (2018) tempers this somewhat, allowing that what matters is what you are *in a position* to know (which is still factive).

## 10 RATIONALITY IN RETROSPECT

### 2.2 The Performance Argument for Internalism About Rationality

The weakness of the challenge posed to the orthodox view by Williamson's anti-luminosity argument is that it doesn't directly challenge the orthodox view at all. It only challenges the classical argument for internalism about rationality. But views in philosophy are not false just because there is one unsound argument in favor of them. So in order to assess what kind of challenge is posed to the orthodox view by anti-luminosity, we need to better assess how central the classical argument is to the core attractions of the orthodox view.

One reason that you might suspect that the classical argument is central to the attractions of the orthodox view is its close connection to the naïve view. Not only is rationality incapable of satisfying the constraints on right blame that are set out in the classical argument if no mental condition is luminous, but rationality cannot be the sort of guide to choice that is envisioned by the naïve view either. Being this sort of guide requires being the sort of thing where we can first determine which way the guide points. But given anti-luminosity, sometimes we are in a better position to know that we satisfy some worldly conditions than we are to know that we satisfy some of the internal conditions of a choice's being irrational.

I conclude that rationality cannot serve its putative role as a prospective guide to choice that is envisioned by the naïve picture. This is simply incompatible with the conclusion of the anti-luminosity argument. But the problem, I want to suggest, lies in the naïve picture—not in the orthodox view of the contrast between rationality and correctness. The point of assessments of rationality, I suggest, is not that they help us to determine what to do—about which they are fallible precisely to the extent that our own mental lives are not perfectly transparent to us—but rather that they are helpful tools of *retrospective* assessment of how well we *performed* in making some choice.

The perspective of retrospection pushes into focus the differences between our information and the facts, because sometimes we learn new information in the meantime, and sometimes we discover that we were mistaken earlier. The intuitive support for the naïve picture comes

from focusing on how this new information affects our retrospective judgment of what we should have done. But it *also* makes salient and interesting the *difference* between the question of what to do and the question of what it makes sense to do in light of our own information. At the time of acting, Bernie is not sitting around thinking about his own beliefs about what is in his glass—he’s just thinking about what is in his glass. Amy is not at the time of choice sitting around thinking about her own mental life, but rather about where the money is. And the poker player is not thinking about what they believe, but rather about what their opponent’s cards might be. Each of them could (imperfectly, of course, given anti-luminosity) think about these things, but they don’t. So insofar as what it is rational for them to do depends on their information, I conjecture, they are not thinking about what it is rational to do either.

In contrast, the *change* in their information when they look back retrospectively *forces* each to distinguish how they would decide *for* their past selves from how they want themselves to decide in the future in similar situations. And each of them will hope that their future self will *do the best with the information it has*. But given anti-luminosity, the best way to do the best with the information that you have, is for you to rely directly on your information—not to rely on information about what information you have. After all, given anti-luminosity, the information that you have about what information you have can come apart from what information you actually have. But fortunately, if facts about what information you have are themselves facts about your psychology—about the inner workings of your own mind—then they can be causally efficacious *directly*, whether or not you believe that they are true or not. In contrast, if facts about what information you have are not internal to you—if they do not supervene on the space inside your skin—then it is hard to see how any designer could calibrate you to be directly sensitive to your information without being mediated by the intermediary of your own judgments about what your information is.

What the foregoing argument illustrates is that rationality is a kind of well-tuned *performance*. When we retrospect on our past choices, that gives us the opportunity to fine-tune ourselves to do better or to maintain our level of performance in similar situations in the future.

## 12 RATIONALITY IN RETROSPECT

Reflecting on the bet in poker where we called and lost, we steel ourselves to make it again next time, knowing that the odds were in our favor despite how it turned out. Reflecting on our trip to the emergency room after sipping gasoline, we steel ourselves not to give up drinks poured by other people in the future, because we know our action did make sense at the time. And reflecting on the chance to have gotten the larger sum of money from the second envelope once it is revealed, we steel ourselves to taking the first one again the next time—or after guessing the third envelope and ending up with nothing, we try to re-calibrate and plan to go with the sure thing or expectedly best choice next time.

When we steel ourselves in any of these ways to make similar choices even though we know that the ones that we did make did not turn out—or try to re-calibrate and plan to make a better choice next time, one more attuned to the information available to us—we are at work fine-tuning our decision-making processes. Because anti-luminosity is just as true for retrospective knowledge of our own mental states as it is for concurrent knowledge of our own mental states, of course, no such fine-tuning can be perfect. But by and large we have enough retrospective knowledge of what information we had at the time of choice that there is ample room to make progress through such fine-tuning. And as we fine-tune ourselves in this way, we improve our performance as rational decision makers.

So putting the argument in this section in a slightly different way, rationality as fine-tuned performance requires rationality to depend on wholly internal states, because fine-tuning can only be sensitive to our internal function. The only way to fine-tune a machine to be sensitive to features of its environment is to provide a causal role for correlates—representations or otherwise—of those environmental features in its own function. But a machine can be sensitive to its *own* state without ever having to go by proxy. So the right way for a decision-making machine to respond to its environment in the best way enabled by its own information is for it to *think about the world*, and since its thoughts about the world are themselves internal states, for those thoughts to do their job in a functional mental economy that is well-tuned to perform as well as possible at making decisions.

Importantly, this does *not* require the decision maker's internal representations to be luminous to it. It does not even require the decision

maker to even be capable of representing their own internal representations or any other aspect of their own mental life. All it requires is that their internal representations be able to be causally efficacious in their own design. And it is this—not the requirement that mental states be luminous so that we can think about our own mental lives in order to make rational decisions—that requires that rationality be an internal state. Only internal states can figure in the kind of fine-tuning appropriate to a well-performing choice maker.

This gives us a defense of the orthodox claim that rationality depends only on the agent's information, independently of whether it is true. But instead of arguing for it from the orthodox view that irrational actions merit blame and assuming the prospective perspective of deliberation, we have argued for it from the role of assessments of rationality from the retrospective perspective. But we can also use this same retrospective perspective on rationality in order to explain why rational actions merit praise and irrational actions merit blame—as also claimed by orthodoxy, and given up by proponents of anti-luminosity who use P2 to reject P1. This is because praise and blame are agent-directed forms of evaluation. When we praise or blame you, we are assessing your performance. But rationality, as we saw, is a measure of your performance. How rational your action is measures how well you performed in doing it—not the luck of your circumstances. In contrast, praise and blame are not suited for aspects of your action that come from the luck of your circumstances rather than from your own contribution.

### 3.1 *De Re* Prospective Corrective

In the last section I argued that rationality is actually the appropriate *retrospective* on choice, and that this vindicates the commitment of the orthodox view of rationality that it depends only on the choice maker's internal state. This constitutes my answer to the challenge from externalism to the effect that there is no reason to think that rationality must be internal. The problem with the challenge is that it assumes, along with some proponents of the orthodox view, that it is part of the orthodox view that rationality must be suited to play the prospective role in

## 14 RATIONALITY IN RETROSPECT

helping to guide choice—and that it can do this only if we are always in a position to know which choices are rational. But I say: rationality is instead the appropriate retrospective perspective on choice. It is what we pay attention to in order to fine-tune our performance and assess whether others' performance is fine-tuned. And in order to play this role, it must depend only on things that can feature directly in fine-tuned performance—on states or processes that are internal.

This argument turned on the claim that part of good performance in making decisions is thinking about the *world* rather than thinking about our own minds. If you are thinking about your own state of mind, then we simply introduce an extra layer of possible error between your thoughts that are active and the features of the world that you are responding to, except in the very special circumstance where it is the state of your own mind that matters objectively. And, in fact, this seems right. As we noted, none of Amy, Bernie, or our poker player is concerned about their own mental lives in deciding what to do. And given the arguments that I have just given, rightly so.

So that leaves us with the question of whether there is any category of normative evaluation of action that is as helpful *prospectively* as rationality is helpful in retrospection—or as helpful prospectively for guiding choice as those who endorse the classical argument for internalism about rationality mistakenly took rationality to be. Such a category of normative assessment must be one that we can apply prospectively such that thinking about what to do and thinking about which choice is it are rationally equivalent, so that settling one question is tantamount to settling the other.

We saw earlier that the three envelope problem is commonly conceived of as an argument that correctness could not play this role. Amy, after all, could reason by cases that choosing envelope 1 is incorrect. So if Amy's reasoning is sound, and I am right that rationality is the appropriate perspective for retrospection, then maybe there is no normative concept at all that can help us prospectively by guiding our choices. Good practical reasoning, we might conclude, is *never* in the first instance reasoning about what one ought to do, nor does learning what one ought to do ever settle such reasoning. And the same goes for any other normative concept—none of them can appropriately play such a

role. If this is right, then the way to reason about what to do is always to pay attention to what you ought to do read *de re*—in the sense of paying attention to the things that *are* the things that you ought to do, rather than paying attention to them under that or any other normative description.

If all of this is right, then it would be grist for the mill of those who say, along with Bernard Williams, that saving your wife because it is the permissible thing to do instead of simply out of recognition that she is drowning is one thought too many. It would lend support to those who say, with Michael Smith (1994), that good character requires concern with doing what is right *de re* but not *de dicto*. And it would support the position of those, who say with Nomy Arpaly (2002) and Julia Markovits (2010), that acting with moral worth does not require acting in knowledge that what one is doing is the right thing to do.<sup>9</sup>

So all of that would, I think, be an interesting and defensible result. And I offer it to those who hesitate to follow me the remainder of the way. Nevertheless, I suspect that it is not the correct conclusion to draw. And that is because, as I will now argue, it is a mistake to think that the three envelope cases support this conclusion. The objective ‘ought’ of correctness—the one that depends on the facts independently of the agent’s information—is, in fact, the appropriate and helpful concept to use from the prospective position of choice. Not only is it helpful in order to decide what to do, but it is rational to do something just in case it is rational to conclude that it is what you ought—in the sense of correctness—to do.

### 3.2 Four Envelopes: Or, Why Correctness Is Helpful in Prospect

In the three envelopes case, we get a case in which what it is rational to do comes apart from what it is correct to do, or from what the agent in question ought to do. We can see that, as observers of the case, because we occupy the perspective of having the correct information. We know

<sup>9</sup> Contrast Sliwa (2016).

## 16 RATIONALITY IN RETROSPECT

that the most money is in the second envelope, and hence that that is the choice that Amy ought to take. And it is the choice that we would advise her to take if given the opportunity. But since the case draws on our judgments about what Amy ought to do, and simultaneously gives us information about in order to draw that judgment, nothing about the case helps us to discriminate whether *which* judgment we make about what Amy ought to do is filtered by *our own* level of information about the case.

We can test this by adapting the three envelope case in order to incorporate the perspectives of agents with different levels of information.<sup>10</sup> The four envelope case is like the three envelope case in that there is an agent, in this case Xiao, who gets to decide which of some number of possibly cash-containing envelopes to select—in this case, four. Xiao knows that the first envelope contains \$4000, and that among the other three envelopes one contains \$5000, one contains \$6000, and one is empty. So since the first envelope is a sure thing for \$4000 and any other envelope has only an expected value of \$3667 or so, Xiao thinks that she ought to take the first envelope. When she says, ‘I ought to take the first envelope’, no one will be surprised.

But in the four envelope case, in addition to Xiao, we add two observers.<sup>11</sup> The first observer is Ying, who knows everything that Xiao knows, but also knows that the \$5000 is in the second envelope. And the second observer is Zach, who knows everything that Ying knows, but also knows that the \$6000 is in the third envelope. Ying, it seems, can reason as follows: the second envelope is a sure thing at \$5000, but the first envelope is only \$4000 and the other two have expected values of only \$3000. So Xiao, clearly, ought to take the second envelope. If he had

<sup>10</sup> I first stated the four envelope case in Schroeder (2018), but it is a modified version of cases from Kolodny and MacFarlane (2010) and Bronfman and Dowell (2016) that removes some of their distracting features and ensures that all three speakers make contrasting ‘ought’ judgments.

<sup>11</sup> According to some perspectival accounts, such as that of Björnsson and Finlay (2010), what Xiao ought to do depends in part on what information she could acquire before the time of choice, and not just on what information she already has. So, to be careful, let’s add the stipulation that neither Ying nor the other observer, Zach, can or will share their knowledge with Xiao before she has to choose.



the opportunity to advise Xiao on her choice, this is what he would advise her to do, and if he is sympathetic to her financial plight and learns that she took the second envelope, he will be more pleased than if he learns that she took one of the other envelopes. So if Ying tells us, ‘she ought to take the second envelope’, no one should be surprised.

Zach, of course, reasoning in the same way, will be led to the conclusion that Xiao ought to take the third envelope. After all, he thinks, it is a sure thing at \$6000, whereas the first envelope nets only \$4000 and the second envelope nets only \$5000—and the last one is empty. Since the third envelope is expectedly best, it is the one that Xiao should take. If given the opportunity to advise Xiao before her choice, that is what Zach would tell her to do, and if he is sympathetic enough to her financial plight, it is the choice that he will be most pleased to learn that Xiao has taken. So if Zach tells us, ‘she ought to take the third envelope’, no one should be surprised.

Now, between Xiao, Ying, and Zach, we have three different and apparently conflicting judgments about what Xiao ought to do. Xiao thinks that she ought to take the first envelope, Ying thinks that she ought to take the second, and Zach thinks that she ought to take the third. Now, it *could* be that they are talking past one another, and that Xiao is really interested in and talking about which choice it is *rational* for her to make, expressible by a ‘subjective’ sense of ‘ought’ that is relative to the agent’s beliefs or evidence, whereas Zach is really interested in and talking about which choice it is *correct* for her to make, expressible by an ‘objective’ sense of ‘ought’ that depends only on the totality of the facts. But that leaves mysterious what Ying is talking about.

Ying cannot be intelligibly interpreted as talking about what Xiao ought subjectively to do, in light of her own information, because he knows full well that Xiao does not know where the \$5000 is. So he knows that in light of her own information, the rational choice is to take the first envelope. He just wouldn’t advise her to do so, given the chance, because he does know where the \$5000 is. Nor can Ying be intelligibly interpreted as talking about what Xiao ought to do in light of all of the facts, because he knows full well that there is an answer as to where the \$6000 is. Rather, Ying’s claim about what Xiao ought to do, it seems, is filtered through *his own* information.

## 18 RATIONALITY IN RETROSPECT

Once we see that Ying's claim about what Xiao ought to do must be filtered through his own information—and not through either Xiao's information or through his estimation of the totality of all of the facts—we can see that Xiao's and Zach's claims can also be understood in this way. Given that Xiao does not know where the \$5000 or \$6000 is, reasoning about what she ought to do *in the very same way as Ying does* should lead her to conclude that she ought to take the first envelope. We do not need to appeal to a different meaning or contextually supplied content of 'ought' in order to make sense of what Xiao says. When she makes a claim as speaker about herself as the agent of the 'ought', it is her role as *speaker* that is relevant, and not her role as agent. That is what it takes in order to make her use of 'ought' continuous with Ying's. Similarly, given that Zach does know where the \$6000 is, reasoning about what Xiao ought to do in the very same way as Xiao and Ying reason should lead him to conclude that she ought to take the third envelope. This is the choice that maximizes expected value relative to *his* information. And it only happens to coincide with the best choice for Xiao overall because he happens to have all of the relevant information—not because he is trying to make a claim that is relative to all of the facts, whatever they are and independently of whether he knows about them.<sup>12</sup>

### 4.1 Deliberation, Advice, and Perspective

If this is how 'ought' judgments work, then they are appropriate both for deliberation and for advice. They are appropriate for advice because they reflect the best judgment available on what to do in the agent's situation, given the information available from the position of the advisor. That is why each of the characters in the four envelope case would advise Xiao differently. Each would be giving her the best advice that they can, given the information available to them. But 'ought' judgments are also helpful from the prospective perspective of deliberation, because when you are

<sup>12</sup> In general, to set up a case with  $n$  distinct levels of information and contrary judgments by  $n$  agents, it is sufficient to set up  $n+1$  envelopes, with  $\$1000 \times (\frac{1}{2}(n-1)n+m)$  in the  $m$ th envelope for envelopes  $1-n$ , and  $\$0$  in the last envelope. If the correct decision rule under uncertainty is not expected value maximization, then these numbers may have to be adjusted.

the agent in question, the most informed advice that you can give the agent is your own.

Of course, as philosophers, we know that there are different ways in which a statement might be claimed to be relative to a speaker's information. It could be that the claim is context-dependent, expressing a different content in different contexts of utterance, depending on who the speaker is and what their information is. For example, it could be that 'I ought to do it' is equivalent to 'it maximizes expected value relative to my own information'. This, of course, would make 'ought' claims inward-looking in the way that I have argued that they are not and cannot be, if 'ought' judgments are to guide decision making from the prospective perspective of choice.

On a different sort of contextualist view, 'I ought to do it' is equivalent to 'it maximizes expected value relative to E', where E is the information available to the speaker, whatever that might be. On this view, 'ought' claims are not themselves inward-looking—they are outward-looking, paying attention to the evidence itself and not to the evidence under the description that it is the speaker's own. Nevertheless, it seems to me that a competent speaker, who knows the truth-conditions of 'I ought to do it', will only be rationally licensed to believe that this is true if she is licensed to believe that it maximizes expected value relative to her own information—that is, roughly speaking, if she is licensed to believe what is sometimes called the 'diagonal proposition' by practitioners of two-dimensional semantics. And so, it seems, this contextualist interpretation will lead rational first-person judgments of what one ought to do to come apart from rational choice precisely in the ways that first-order beliefs are neither luminous nor transparent.

A different way of interpreting the idea that 'ought' judgments are sensitive to the speaker's information, of course, is the idea that they express *relatively true* propositions that are evaluable for truth only relative to what MacFarlane calls a context of *assessment*. This, of course, is MacFarlane's own view of 'ought' judgments, as developed in his book and in joint work with Niko Kolodny—as part of which they give cases intended to raise the same issues as four envelope cases.<sup>13</sup> But according

<sup>13</sup> Kolodny and MacFarlane (2010); MacFarlane (2014).

## 20 RATIONALITY IN RETROSPECT

to MacFarlane, to believe an assessment-sensitive proposition is to believe it to be true relative to one's own context of assessment. So if the feature of contexts of assessment that matters for the assessment-sensitive truth of 'ought' claims is the information of the speaker of that context of assessment, then again believing that you ought to do it requires believing something about your own information—which again brings us into trouble with the luminosity and transparency of our own mental lives. On none of these views are first-personal objective 'ought' judgments more helpful for deliberation than first-person 'rational' judgments.

The expressivist interpretation of the speaker-sensitivity of 'ought' judgments, however, yields a quite different result. According to the expressivist interpretation of the speaker-sensitivity of information, believing that it might happen is not believing something about your own mental life, but just having a positive credence that it will happen. Similarly, believing that it will probably happen is not believing that it is probable according to your own credences, but just having a credence on which it is probable. Likewise, believing that if it happens now, it probably won't happen again later is not believing that you have a high conditional credence that it won't happen later, conditional on it's happening now, but just having such a high conditional credence. And similarly, believing that you ought to do it is not believing something about the expected value relative to your own information, but simply believing something about the expected value *simpliciter*, where probability talk is given the expressivist interpretation.<sup>14</sup>

So, on the expressivist interpretation, claims about what someone ought to do always express our own epistemic perspective or information state. We believe these claims in virtue of *being in* such an information state, rather than in virtue of believing that we or anyone else is in such an information state. And that is why they can be directly effective in helping us, from the prospective perspective, decide what to do. The rational answers to the question of what to do and the question of what one ought to do are the same, on this view, because both are directly

<sup>14</sup> Compare Yalcin (2007, 2012); Moss (2016); Lennertz (2021).

shaped by your information—*without* needing this to be mediated through beliefs or representations about what your information is.<sup>15</sup>

## 4.2 Catching Our Breath

The suggestion that I have been offering in this chapter is a bit dizzying. I certainly find it so, and I am the one offering it. In a way, it upsets everything that I thought that I knew, about both choice and what we ought to do. But in another way, it is deeply conservative. I'll close with a few thoughts about each of these points, in reverse order.

In Sections 3.2 and 4.1, I have been arguing that 'ought' does not work at all in either of the ways that many philosophers have presumed. It is not sensitive in any way to the agent's information, as 'rational' is. And it does not care about all of the facts, when the speaker is uninformed about some of those facts. It is, instead, expressive of the speaker's own state of information, incomplete as it might be. And this is precisely what, I have argued, makes it suitable both for advice and for deliberation. Any 'ought' that depends instead on the agent's information *qua* agent will *ipso facto* be a less helpful guide to deliberation, because from the prospective perspective of deliberation our imperfect grasp of our own mental lives only creates an additional layer by which such judgments could lead us astray. Far better if our choices are informed directly by our information than if they are informed by our information about our information.

I said at the beginning that my aim in this chapter was to vindicate the orthodox view, but in a surprising way. According to the orthodox view, recall, what it is rational for someone to do depends on the information that she has, and someone is to be praised for doing what is rational and

<sup>15</sup> I have not, of course, specified an exact way in which 'ought' judgments are expressive of information, and it is important to note that there are very important choice points along the way. One choice concerns what kind of decision-rule to implement. But another concerns whether 'ought' simply imports the decision-rule directly relative to the speaker's own information, or instead reflects which advice she should give the agent. Choice points like these will be very important for evaluating the implications of the view for cases in which observers have less information than the agent.

## 22 RATIONALITY IN RETROSPECT

blamed for its opposite. I have vindicated these claims in Section 2.2, not by showing that our mental lives are transparent to us, but by showing why rationality is a measure of well-tuned performance. If you do not do what is rational, that is a measure of *your* failure of performance—not of your environment. And that—rather than your being in a position to know how to avoid it—is why it merits a kind of blame. In contrast, if you *do* do what is rational, then that is a measure of your *well-tuned* performance. And so because it is a measure of *you*, rather than of your environment, you are the one who deserves a kind of praise. Finally, as we saw before, in order to be an assessment of your performance, your rationality must depend only on features internal to you, because only such features can be tuned toward better performance.

So much for the vindication of the rationality component of the orthodox view. But I can also vindicate the correctness component of the orthodox view—so long as we are careful not to over-interpret that component of orthodoxy. According to the orthodox view, recall, what it is correct for someone to do depends on the facts, independent of what they believe, and it is appropriate to congratulate someone for doing what is correct, and to commiserate with them for its opposite.

But now consider the hypothesis that what is correct goes with what someone *ought* to do, as I have explored in Sections 3.2 and 4.1. What Xiao ought to do, all three of Xiao, Ying, and Zach will agree, depends on the facts, independently of what Xiao believes about them. This is clear, of course, in the case of Zach, which is the most like our own case as observers when the original three envelope case is described to us. But it is also clear in the case of Ying, who knows full well that Xiao is not aware of where the \$5000 is, but thinks that matters only for what it is rational for her to do, and not for what she ought to do. And though it is less obvious, it is also true for the case of Xiao, who should admit that it is really neither here nor there what she believes about where the \$4000 is, but only relevant that it is in fact (as she believes) in the first envelope. In all three cases, then, our speakers should agree with the orthodox view that what Xiao ought to do depends on the facts independently of what Xiao believes about them. It is no wonder that we think this about correctness and likewise about what people ought to do, of course, because we as theorists have nearly invariably been in the situation of

Zach when confronted with cases intended to distinguish rationality from correctness.

The other feature of the orthodox view is that it prescribes congratulations for the correct choice and commiseration for the incorrect one. But this, too, is vindicated by the perspective-expressive use of ‘ought’ that we have been exploring. Suppose that instead of maximizing expected utility, Xiao chooses the second envelope at random, or because her favorite number is two. And suppose that while they wait to see what is in the envelope, Ying gets to talk to Xiao. Surely it makes sense for him to congratulate her on doing better than \$4000! Of course, once both learn that the \$6000 was in the third envelope, both of their information will change. Ying and Xiao will both come to believe that she ought to have taken the third envelope. And, of course, this change in their perspective is part of what facilitates the intuitive illusion that correctness is appropriate to the perspective of retrospection. But once they do, it then makes sense for Ying to stop congratulating Xiao and begin to commiserate with her—as in ‘Darn—you should have taken that one. Oh, well—you do the best that you can.’

So flipping our orientation to understand rationality as the mode of assessment appropriate to retrospection and correctness as the mode of assessment appropriate to prospection turns out—with some substantive commitments along the way—to provide a surprisingly thorough vindication of the orthodox view about the relationship between rationality and correctness. The one thing that it does *not* vindicate, that might be thought by many to be part of the orthodox view, is the idea that correctness depends on the totality of *all* of the facts. That idea, of course, is the one that makes it look like correctness *cannot* serve as a guide to choice in three envelope cases. And we have had to reject that idea along the way.

But we have also generated an error theory for why people could have come to think that this is true. That is because when we consider philosophical thought experiments, we generally take ourselves to be omniscient about the cases—after all, we can stipulate any answer to the details that we like. So when we make judgments about what the agents in those cases ought to do, even though we are making them relative to our own information, we are like Zach—unable to distinguish

## 24 RATIONALITY IN RETROSPECT

what holds relative to our own information from what holds in light of all of the facts. And in other work, I have gone even farther, and argued that the idea that correctness corresponds to what we ought to do in light of *all* of the facts is even closer to the truth.<sup>16</sup>

All of this amounts, I claim, to evidence that my conclusions in this chapter are surprisingly conservative—they really do vindicate the orthodox view at its core; they really do make sense of what we have been talking about all along, about rationality and correctness; and they really do vindicate the orthodox view against the challenges that we encountered earlier. But I must admit that the way that I have gotten to these conclusions is unsettling. The most unsettling of all, I conjecture, are my theses that ‘ought’ is always expressive of the speaker’s own information and that talk about what is correct in light of the facts and independent of the agent’s information is nothing over and above talk about what she ought to do, so construed.

If I am right, then when Derek Parfit used examples to introduce what he calls ‘ought’ in the ‘fact-relative’ sense, he failed entirely to do so. What he did, instead, was to illustrate cases in which what someone ought to do depends on facts of which they are unaware, and imposed the false theory that what they ought to do in that sense thereby depends on all of the facts—a false theory whose allure is enhanced by the fact that as we learn more of the facts, our perspective shifts to accommodate them.

But if there is no fact-relative sense of ‘ought’, and talk about which choices are correct is just talk about what people ought to do, then we are going to need to go back over a great deal of our moral theorizing in order to unpack how it has been filtered by the assumption that what matters for rights, or duties, or the like must be a matter either of the totality of the facts or of the information available to the parties. If any of ethics is to tell us anything about what we ought to do, then there is much that we must comb over in order to be clear on where it has been shaped by this misunderstanding. That does make me dizzy. But at the same time, it brings all of ethics into a kind of perspective.<sup>17</sup>

<sup>16</sup> Schroeder (2018).

<sup>17</sup> Special thanks to an audience of a very early predecessor of this chapter at Shandong University in May 2018, and to audiences of a mature version of the chapter in fall 2020 at Boston University, the University of Vienna, the Madison Metaethics Workshop, and the Frontiers of Contemporary Metaethics conference at Nanjing Normal University.



## References

- Arpaly, Nomy (2002). ‘Moral Worth.’ *Journal of Philosophy* 99(5): 223–45.
- Berker, Selim (2008). ‘Luminosity Regained.’ *Philosophers’ Imprint* 8: 1–22.
- Björnsson, Gunnar and Stephen Finlay (2010). ‘Metaethical Contextualism Defended.’ *Ethics* 121(1): 7–36.
- Blackson, Thomas (2007). ‘On Williamson’s Argument for his Anti-Luminosity Argument.’ *Philosophy and Phenomenological Research* 74(2): 397–405.
- Bronfman, Aaron and Janice Dowell (2016). ‘Contextualism About Deontic Conditionals.’ In Nate Charlow and Matthew Chrisman, eds., *Deontic Modals*. Oxford: Oxford University Press, chapter 4.
- Brueckner, Anthony (2002). ‘Williamson’s Anti-Luminosity Argument.’ *Philosophical Studies* 110(3): 285–93.
- Gibbons, John (2009). ‘You Gotta Do What You Gotta Do.’ *Noûs* 43(1): 157–77.
- Gibbons, John (2013). *The Norm of Belief*. Oxford: Oxford University Press.
- Kiesewetter, Benjamin (2016). ‘You Ought to  $\varphi$  only if You May Believe that You Ought to  $\varphi$ .’ *Philosophical Quarterly* 66(265): 760–82.
- Kiesewetter, Benjamin (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.
- Kolodny, Niko and John MacFarlane (2010). ‘Ifs and Oughts.’ *The Journal of Philosophy* 107(3): 115–43.
- Lennertz, Benjamin (2021). ‘Noncognitivism and the Frege–Geach Problem in Formal Epistemology.’ *Philosophy and Phenomenological Research* 102(1): 184–208.
- Lord, Errol (2010). ‘Having Reasons and the Factoring Account.’ *Philosophical Studies* 149(3): 283–96.
- Lord, Errol (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.
- MacFarlane, John (2014). *Assessment-Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.
- Markovits, Julia (2010). ‘Acting for Right Reasons.’ *Philosophical Review* 119(2): 201–42.
- Moss, Sarah (2016). *Probabilistic Knowledge*. Oxford: Oxford University Press.

## 26 RATIONALITY IN RETROSPECT

- Ramachandran, Murali (2009). 'Anti-Luminosity: Four Unsuccessful Strategies.' *Australasian Journal of Philosophy* 87(4): 659–73.
- Regan, Donald (1980). *Utilitarianism and Cooperation*. Oxford: Oxford University Press.
- Ross, Jacob (2006). 'Acceptance and Practical Reason.' PhD dissertation, Rutgers University.
- Ross, Jacob (2012). 'Rationality, Normativity, and Commitment.' In Russ Shafer-Landau, ed., *Oxford Studies in Metaethics, Vol. 7*. Oxford: Oxford University Press, pp. 138–81.
- Schroeder, Mark (2018). 'Getting Perspective on Objective Reasons.' *Ethics* 128(2): 289–319.
- Sliwa, Paulina (2016). 'Moral Worth and Moral Knowledge.' *Philosophy and Phenomenological Research* 93(2): 393–418.
- Smith, Michael (1994). *The Moral Problem*. Oxford: Basil Blackwell.
- Srinivasan, Amia (2015). 'Are We Luminous?' *Philosophy and Phenomenological Research* 90(2): 294–319.
- Williams, Bernard (1981). 'Internal and External Reasons.' In his *Moral Luck*. Cambridge: Cambridge University Press, pp. 101–13.
- Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
- Yalcin, Seth (2007). 'Epistemic Modals.' *Mind* 116(4): 983–1026.
- Yalcin, Seth (2012). 'Bayesian Expressivism.' *Proceedings of the Aristotelian Society* 112: 123–60.