

WILEY

What to Do When You Don't Know What to Do When You Don't Know What to Do...

Author(s): Andrew Sepielli

Source: *Noûs*, September 2014, Vol. 48, No. 3 (September 2014), pp. 521-544

Published by: Wiley

Stable URL: <https://www.jstor.org/stable/43828886>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Wiley is collaborating with JSTOR to digitize, preserve and extend access to *Noûs*

*What to Do When You Don't Know What to Do When You Don't Know What to Do . . .*¹

ANDREW SEPIELLI

There's been renewed interest in the question of what it's rational to do in the face of normative uncertainty—uncertainty among moral theories, claims about what to do in particular cases, and even accounts of rationality itself.² But some fascinating problems have been raised not for any particular answer to this question, but to the very project of trying to answer this question. One of these problems invokes agents who've tried to answer the question for themselves, but have arrived at answers that we might regard as mistaken. Consider:

Solomon must determine the fate of the murderer Norman. Solomon knows that punishing Norman will achieve none of the “forward-looking” aims of punishment, and wonders what to do. On one hand, he thinks retribution is probably a sufficient, if weak, ground for punishment. But he thinks it might be no ground whatsoever, in which case jailing someone for sheer retribution would be horribly wrong. It would be, essentially, locking a person in a cage for no good reason. In desperation, he consults a book about rationality under moral uncertainty.

It reads: “. . . if an agent thinks it is more probable than not that doing A is better than doing B, it may still be more rational to do B, if the difference in moral value between doing B and doing A is much greater, on the supposition that B is better, than the difference between doing A and doing B is, on the supposition that A is better. In other words, it can be rational to ‘hedge your bets’ in cases of moral uncertainty, and act on views that are less probable, but have more moral value at stake.” On this view, let us assume, it would be most rational for Solomon to free Norman.

But Solomon just doesn't buy it. He thinks this author's “moral hedging” idea is crazy. He comes to believe instead that it is rational simply to do the action that is most probably right. On this view, Solomon rationally ought to jail Norman. And so he does just that.

The author of the moral uncertainty book is told about all of this, and is asked about the rationality of Solomon's action.

What should our author say?

Perhaps she should say it would have been more rational for Solomon to free Norman. After all, that's what the theory in her book implied. If Solomon didn't want to believe that theory, that's fine, but it doesn't make the theory false, any more than someone's disbelief renders a scientific theory false.

This answer is not entirely convincing. Rationality, as I shall understand the notion, is supposed to be sensitive to the subject's perspective. So how can the

author say what it's rational for Solomon to do is utterly independent of his perspective about what to do in the face of moral uncertainty? And imagine yourself in Solomon's place. If you jailed Norman, that would feel natural, as though the action flowed ineluctably from your conviction about what to do. But if you freed Norman, your action would seem alien to you. After all, you've settled on one action, but then performed another. It doesn't seem right, in light of this, to call freeing Norman the rational action.

Perhaps she should say that it was indeed more rational for Solomon to jail Norman. Rationality is about doing what's right from your perspective, and sending Norman to jail is right from Solomon's perspective.

This answer is odd as well. Rationality as we're going to define it is subjective, sure, but this makes it *too* subjective. It's as if Frege had said, "The laws of logic are the normative 'laws of thought'³ ... that is, unless you don't believe 'em, in which case, y'know, think whatever."⁴ It also turns the task of trying to find out what to do under moral uncertainty into a bit of a farce. For it makes what the agent should do dependent upon what the agent *believes* he should do. Knowing this, an agent should just settle on an answer by fiat, confident that it will verify itself.

Perhaps our author should say that in one way it was more rational for Solomon to jail Norman, but that in another way it would have been more rational for Solomon to free him. Both of the previous answers had their merits, so why not combine them?

This answer also seems unsatisfactory. First, both of the previous answers had their *flaws*, so that's a reason *not* to combine them. Second, the idea that it's both more rational to do one thing and more rational to do the other is of dubious coherence. If it's not like saying George Bush is both taller than and shorter than Barack Obama, we'd have to hear how. Third, even if this idea can be rendered coherent, it seems like there should be some kind of *overall* assessment of rationality. Otherwise, the author's not exactly offering helpful advice, is she?

Seeing the author backed into a corner like this, we might be tempted to conclude that she shouldn't have written the book in the first place—that she took a wrong turn in even trying to give a theory of rationality under moral uncertainty. I would find this conclusion very dispiriting. I don't know about you, but I'm uncertain about lots of moral questions, and important ones at that—which economic systems are fair, when if ever wars are just, which if any moral theory is true, and, yes, whether retribution is a sufficient ground for punishment. The question of what to do in the face of this uncertainty seems to cry out for an answer—"cry out" in the sense that, unless there exists such an answer, my actions can be nothing other than stabs-in-the-dark, leaps of faith, guesses. Which is all to say I share our imagined author's ambitions.

So I shall claim there is something perfectly satisfactory and even illuminating for the author to say about cases like Solomon's. My view is that: (a) there are two very different conceptions of rationality at play in the intuitions above: "perspectival" and "systemic"; (b) it would be rational in the perspectival sense for Solomon to jail Norman; and (c) it is quite plausible there is no one thing it is systemically rational to do; rather, there may be incommensurable "orders" of systemic rationality

and these orders may disagree about what to do. Put simply, my view is that the second answer above is right if we're talking about perspectival rationality, and that something like the third answer may be right if we're talking about systemic rationality. Of course, I will try to convince you that the problems cited for these answers can be surmounted as they are combined into a single, hybrid view.

What we say about Solomon will be applicable to other cases as well. It will be applicable to the case of someone who is uncertain about what to do under normative uncertainty (rather than, like Solomon, outright mistaken about it). Indeed, it will be applicable to someone who is beset by even higher levels of normative uncertainty—uncertainty about what to do under uncertainty about what to do under uncertainty about what to do, etc.

More generally, it will be applicable to any case in which a person is either uncertain or outright mistaken regarding a putative rule obedience to which is required by rationality. And lots of cases are like this. They arise in theoretical as well as in practical philosophy. At the end of the paper, we'll briefly survey how the framework of this paper might be applied to two theoretical matters: the rationality of moderating one's beliefs in the face of peer disagreement *when one disagrees with a peer about whether to moderate one's beliefs in the face of peer disagreement*; and the rationality of believing the conclusion of a competently executed deduction *when one doubts or disbelieves that one performed a competently executed deduction*.

1. Some Conceptual Apparatus

I'll need to introduce some concepts and distinctions before my view can be articulated and defended. These are: 1) the concept of rationality, 2) the distinction between perspectival and systemic conceptions of rationality, 3) the distinction between local and global rationality, and 4) the concept of orders of rationality.

1.1 Rationality

"**Rationality**" is used by different people to express different concepts. In keeping with a prominent strand in the literature, I will use it to express a purely "subjective" normative notion such that it is a conceptual truth that what is rational for someone depends only upon her mental states.⁵ Sometimes I will treat rationality as a gradable feature; an action or a revision-of-mind may be less rational than some other actions but also more rational than some others.

1.2 Perspectival and Systemic Conceptions of Rationality

As I suggested in the introduction, there seem to be two very different conceptions of rationality at play in our intuitions about Solomon's case. On the **perspectival** conception, the fully rational action or revision-of-mind must (a) appropriately reflect the mental states that comprise an agent's or thinker's perspective, and (b) make sense or be intelligible from that perspective. There is a particular sort of intentional explanation that we arrive at by taking up another's perspective and ascertaining whether the putative explanandum makes sense from that perspective. In the relevant literature, explanations of this sort are sometimes called "reasons

explanations” or “rationalizations”. In light of the foregoing, I think it would add precision to call them “*perspectival* rationalizations”. On the *systemic* conception of rationality, by contrast, the fully rational action or revision-of-mind must only appropriately reflect the mental states that comprise an agent’s or thinker’s intentional system; it needn’t make sense or be intelligible from her perspective.

The viability of this contrast relies, of course, on the viability of the contrast between being appropriately reflective of an intentional system, and making sense from a perspective. This contrast will certainly not seem viable if we think of a person’s perspective as identical with her entire intentional system, and making-sense-from as the same thing as appropriately reflecting.

But this is not what I have in mind. A person’s perspective at a time is comprised of all and only her mental states that are phenomenally conscious at that time. On the prevailing views of the mental, these will be but a subset of the mental states in her intentional system at that time. So the supervenience base for systemic rationality is much wider than that for perspectival rationality.

And while the idea of intelligibility from one’s own perspective is to be understood in terms of a constitutive phenomenology, the concept of appropriately reflecting a set of mental states is a purely normative one. Systemic (ir)rationality is no more characterized by a constitutive “feel” than (il)legality is.

Some would characterize the phenomenology of sense-making or intelligibility as one of “guidedness” or “fluidity”. The perspectivally rational action, in that case, must among other things feel guided or fluid. But as Wittgenstein (among others) suggests, there may be no phenomenology of guidedness. Deep, deliberate reading feels one way, and fast, natural reading another, but in both cases the reader is being guided by the text.⁶

Rather, the phenomenology in the neighborhood seems to be one of *unguidedness*. When I am afflicted with conscious uncertainty about how to proceed, my prospective actions or revisions-of-mind strike me as leaps of faith; I feel like I simply don’t have enough to go on in acting or thinking.⁷ I would want to say, then, that an action is perspectivally *irrational* if it has *this* phenomenology; to be perspectivally rational, an action must, if it is consciously performed, *lack* the phenomenology of being a leap of faith.

For the purposes of this paper, though, it does not matter whether we characterize perspectival rationality in part by a phenomenology of guidedness or perspectival *irrationality* in part by a phenomenology of *unguidedness*. That the perspectival conception is characterized partly in terms of phenomenology at all, and the systemic conception is not, explains how there could be actions which are rational on the latter conception but not the former. For an agent with a set of mental states that a prospective action would appropriately reflect may also have a phenomenally conscious state that is either sufficient to render all prospective actions unintelligible, or insufficient to render any of them intelligible, from his perspective. Before he settles on a conclusion about what to do under uncertainty, Solomon seems to be just such an agent; and if the book’s claims about rationality are correct, freeing Norman is just such an action. According to the author, this action is appropriate in light of Solomon’s uncertainty. But it is not at this time intelligible from his

perspective. It feels, as many of his candidate actions do, as though it would be a leap of faith. That is what spurs him to consult the book in the first place!⁸

There is also an explanation for how there could be actions that are perspectively rational but not systemically rational. An action that makes sense from a person's perspective may be inappropriate in light of the agent's other mental states. If the book's claims about rationality are correct, then it's plausible that *jailing* Norman is just such an action. Solomon's belief that it is the right thing to do renders it perspectively rational, but according to the author, this action is inappropriate in light of Solomon's credences in the different theories of punishment.

1.3 Global and Local Rationality

We noted that systemic rationality seems to have a very broad supervenience base. But how, in action and thought, do we respond appropriately to that base's constituents? For example, what is it rational for a thinker to believe regarding some matter when her other beliefs that bear on that matter are inconsistent? What should she do in light of her beliefs about what to do *given her other beliefs*? What if she believes that other beliefs of hers are mistaken? What if she has mistaken beliefs about which other beliefs she has? These are very difficult questions. But I hope to make some progress on at least the ones that are relevant to our assessment of Solomon and those in structurally similar situations. To do that, we'll need the distinction between global and local systemic rationality, and the subsidiary concept of orders of systemic rationality.

It is common in the literature on rationality to distinguish between **global** and **local rationality**.⁹ Global rationality is relative to all of the mental states in an intentional system, while local rationality is relative to only a subset of these. There are different ways of cashing out "is relative to". We might interpret it as "supervenes upon". We might interpret it as "is (non-causally) explained by". The differences between these two interpretations will end up being important in some contexts,¹⁰ but we can get by in this paper without settling this question.

Sometimes our assessments of rationality are of global rationality, as when we ask what action a fully rational agent would perform, or what advice she would give. But sometimes our assessments of rationality are of local rationality, as when we say that if one believes one ought to do A, one is rationally required to intend or desire to do A, or as when we say that if one believes that the evidence conclusively supports P, one is rationally required not to disbelieve P.¹¹

There is a further issue about whether local rationality is merely *prima facie* rationality, or whether it is *ultima facie* rationality just like global rationality, but restricted in its jurisdiction to a smaller set of items. It is difficult to see how we might resolve this dispute, but fortunately, what I say in this paper should go forward equally well on either construal of the relationship between local and global requirements. As we know from reading W.D. Ross, the *prima facie* of a duty or requirement doesn't mean it is theoretically dispensable—that we can simply talk about *ultima facie* duties and requirements without distortion or loss of information.¹²

With this distinction in hand, I can already provide you a glimpse of what the author should say about the systemic rationality of Solomon's actions: it would be locally systemically rational relative to some of his mental states for him to jail Norman, and locally systemically rational relative to some of his other mental states for him to free Norman.¹³

1.4 Orders of Rationality

Suppose action A is locally systemically rational relative to one set of mental states, and action B is locally rational relative to another set. There are different relations that local rationality relative to the first set may bear to local rationality relative to the second. Most of these are of little theoretical import. But among the exceptions is the "higher-order than" relation. Rationality relative to some mental states may be higher-order than rationality relative to other mental states. The distinction among the orders of rationality will be crucial for saying the right things about systemic rationality in a case like Solomon's.

As we'll see, correctly spelling out the idea of higher-order rationality is actually pretty difficult, but it isn't hard to get an intuitive grip on it. Suppose that Athena is uncertain among moral theories T1, T2, and T3, and faces a situation where they disagree about what to do. She asks herself, "What should I do given that there's some chance T1 is right, some chance T2 is right, and some chance T3 is right?" Insofar as her subsequent behavior accords with the *correct* answer to this question, she has done the **first-order rational** thing. Our author might say, then, that if Solomon had only acted in accordance with her advice, he'd have done the first-order rational thing. Or, more familiarly: *akrasia* is precisely a violation of first-order rationality.

But now suppose that Athena is uncertain regarding the answer to her question. She has some credence in each of theories T4, T5, and T6 as providing the correct answer. So she steps back even further: "What should I do given that there's some chance T4 is right, some chance T5 is right, and some chance T6 is right?" If she acts in accordance with the correct answer to this question, she has done the **second-order rational** thing. And if she is uncertain regarding whether T7, T8, or T9 is the correct answer to this latest question, and asks herself what she should do given that there's some chance each of them is right, she will have done the **third-order rational** thing if she acts in accordance with the correct answer.

In thinking about the above, it is important to keep in mind the distinction between acting in accordance with a rule, and following the rule. If the correct rule of, say, third-order rationality says I ought to do A, I am third-order rational insofar as I do A. I needn't adopt a belief regarding this rule, or be part of a community the members of which are disposed to act in accordance with the rule, or whatever else following a rule might require (although perhaps following the rule will make it less "accidental" that I'll act in accordance with it).

Now, Athena, unlike Solomon, was beset by **iterated normative uncertainty**. She was uncertain not only about what to do, but also about what to do when she didn't know what to do, and so on. She was useful, then, for illustrating higher and higher orders of rationality. For only uncertainty forces an agent to keep stepping

back and asking what to do. Certainty in response to any normative question tends to halt this process. But the framework of multiple orders of rationality is just as easily applied to Solomon as to Athena. If what our author said in her book was correct, then it is first-order rational for Solomon to free Norman. For that is the action that accords with the correct rule about what to do given Solomon's uncertainty regarding retributivism. But once Solomon reflects on the matter for himself, and becomes certain of the competing, incorrect rule, it becomes second-order rational—rational relative to that certainty—to jail Norman.

In both of these cases, bringing in orders of rationality helps us to accommodate two of the intuitions registered in the introduction—first, that rationality imposes standards of correctness for a person that are independent of whether he acknowledges those standards or believes he's met them; but second, that rationality takes into account a person's perspective, even when that perspective is the product of doubt about the aforementioned standards or about one's having met them. Put simply, rationality regimented by orders is subjective, but not too subjective.

That should give you a feel for how orders of rationality work. Since this idea is not yet commonplace, though, it would be preferable if we could also give a general expression of it rather than simply illustrating it through examples. Let's start by considering the following characterization:

C1: What it's $N+1$ st order rational to do depends upon the agent's beliefs or credences whose contents are propositions about what it's N th order rational for her to do (unless $N = 0$, in which case it depends on the agent's beliefs or credences regarding objective normativity).

C1 may seem almost painfully obvious, but I think it is not quite right. It's worth distinguishing two questions we might ask about this characterization: 1) Is what C1 calls "rationality" properly so-called? That is, does rationality of *any* sort, class, order, phylum, etc. depend upon the agent's beliefs or credences *regarding rationality?*; and 2) Is C1 a proper expression of the as-yet inchoate notion of orders of rationality that I tried to illustrate using Solomon and Athena? It is possible for the answer to the first question to be "yes" and for the second to be "no". The characterization might be capturing some notion of rationality, but not quite the notion that I'm after.

I'm not sure what the answer to the first question is, although I want to briefly show why both the affirmative and negative answers have some appeal. But it's the second question that's ultimately more germane to the aims of this paper, and the answer to this question is "no", although it will take some work to see why.

It would help with the first question to consider a case. Suppose someone has a full belief or certainty that it would be N th order irrational to do A. We may be tempted to say that this would *make* his doing A irrational. After all, in judging that it would be irrational to do A, the agent is judging that doing so would constitute improper functioning—that doing A would be inappropriate given the mental states in light of which N th order rationality is being assessed. And surely, we might think, doing something that you regard as constituting improper functioning is itself improper functioning.

But there are, on the other hand, grounds to suppose that an agent's doing what he judges to be irrational would not itself be irrational. As others have noted, to judge that an action is rational is not to favor doing it. We would not, for example, advise an "ideally coherent Caligula" to do as he did simply because it fit well with his attitudes. And in tallying up the pros and cons of my own prospective actions, it is not at all obvious that "it would be rational" should find its way onto the board under "it would make my child's day" and across from "it would lend support to the junta". (If it did, then any *kratic* action would automatically get an extra "pro"—in effect allowing agents to "bootstrap" reasons into existence.) But if rationality is not, as some put it, "robustly normative" in a way that is relevant to advice or to pros and cons, then judgments of irrationality are not condemnatory in quite the same way that, say, judgments of immorality are. If you're not doing what you condemn by doing what you *think* is irrational, though, we might question whether it really is irrational.

Even if we answer this first question in the affirmative, though, this does not mean that C1 captures the intuitive idea of higher-order rationality. And indeed, I think it does not capture it.

As I argue elsewhere, we can understand "subjective" normative concepts like that of rationality in terms of the primordial action-theoretic concept of *trying*. On such an understanding, the rationality of an action is a measure of the quality of the action *qua* try at doing what the agent has objective reason to do.¹⁴ Specifically, the most rational action will be the action that is the best try for the agent at doing what she has objective reason to do. Of course, good tries may fail. One bases one's tries on one's beliefs or credences; if those are off-target, then one's best tries will be as well. But best tries are the optimal employments of one's beliefs or credences.

The idea of orders of rationality is crucial to filling out this "trying" understanding of rationality. That is because, for agents like Solomon and Athena, who have "stepped back" in the manner described and thus have multiple levels of belief/credence about what to do, different actions will be count as "the best try" relative to different of those levels. Consider Solomon. Given his uncertainty among the moral theories, it seems natural to describe as his best try at doing what's morally valuable whichever action is most rational relative to this uncertainty. If the author of the book is right, this action is freeing Norman. But once he himself forms the belief that, given the probabilities of retributivism, etc., he ought to jail Norman, it seems like the best try relative to *this* belief is to jail him. Without bringing in orders of rationality, our conception of rational action as good trying would be too simple. We'd be unable to acknowledge that the action that optimally employs one level of belief/credence in one's quest to do what is right may not be the action that optimally employs another level of belief/credence toward the same end.

Given this view of what rational orders are about, C1 will succeed insofar as the action which it designates as most rational on each order really would be the agent's best try, given her beliefs or credences to which this order is relative, at doing what she has objective reason to do. By these lights, it seems to me that, generally, C1 does not succeed.¹⁵

For C1 says that higher orders of rationality depend on the agent's beliefs or credences about lower orders of rationality. And rationality, we said, is about what is appropriate given one's beliefs or credences. But whether my action is a good try at doing what I have objective reason to do does not depend on whether, as I see things, that action reflects my own beliefs. Rather, whether my action is a good try towards this end depends on whether, as I see things, that action would be appropriate *given the way the world is or likely is*—that is, appropriate given the sorts of things that make beliefs and credences accurate. A good trier cares about what is appropriate given his own beliefs/credences only because there can be no “gap” from his point of view between these and the way the world is or likely is, and therefore no divergence between what is appropriate in light of these beliefs/credences and what is appropriate in light of those worldly features that render them accurate. Since an account of rationality is an account of good trying, such an account should advert to what is of fundamental rather than derivative concern to the good trier.

It seems, then, that we ought to replace C1 with:

C2: What it's N+1st order rational to do depends upon the agent's beliefs or credences whose contents are propositions about a kind of value, V, where V depends upon the features of the world that render *accurate* the beliefs or credences upon which Nth order rationality depends (unless N = 1, in which case it depends on the agent's beliefs or credences regarding objective normativity).

1.5 Accurate Credences?

But now we're faced with another problem. We know what makes a belief that P accurate. It's that P. But what makes a credence that P accurate? Not a subjective probability that P, for subjective probabilities just *are* credences. And not an objective probability, for this would imply that people who refuse to assign such probabilities (other than zero and 1) to fundamental normative propositions cannot be in situations like Solomon's or Athena's, and cannot exhibit higher-order rational behavior.¹⁶

What about an evidential probability that P? This is tempting, but again, it seems very plausible that one could be in a position like Solomon's despite refusing to assign such probabilities (other than zero and 1) to fundamental norms. Additionally: The sort of fact that makes a credence accurate should stand to that credence in the same way the fact that P stands to a full belief in P. And facts about evidential probabilities don't do the trick. There is a difference between a mental state's being accurate, and its being evidentially supported. A belief that P is accurate if P, and evidentially supported if the evidence supports P. A credence of .2 is evidentially supported if the evidential probability of P is .2; but we've not yet discovered when it is *accurate*.

This last admonishment make lead one to doubt whether a credence of neither zero nor 1 can ever be truly accurate. For suppose that P. Even a thinker whose credence in P is very high (but less than 1) will, if she is coherent, also have *some* credence in $\sim P$, and to that extent fail to get the world right.

Be that as it may, I think there's something that such a thinker *is* getting exactly right. To see what, consider again how P is related to the belief that P. One relationship is that *the proposition is what's stood for by the expression of the belief*. I express the belief with "P", and "P" stands for P. "Expression" here should be understood broadly: I express the belief that P not only when I say "P" out loud; I do it in inner speech when I consciously deliberate in my head. (A German colleague told me he philosophizes *in German*. Despite being from Philadelphia, I sometimes find it useful to philosophize in BBC English) Arguably, I do it in "mentalese" even when my mind churns away *unconsciously*. Even if we think that mental content is determined publicly, expression is something I can do privately.

I propose, then, that we identify the accuracy condition of a doxastic state with the proposition that is stood for by the expression of that state. In other work, I argue that we express credences using what I call "minimal probability" (or "MP") statements, and that the semantic features of those expressions are explained, in the standard expressivist manner, by the inferential features of the states they're used to express.¹⁷ Of course, the ordinary person does not use the term "minimal probability", any more than she uses the terms "subjective probability" or "objective probability". Rather, he simply speaks of "chance" and "likelihood". But when the function of such speech is to express his uncertainty—rather than, say, to express a full belief about evidence, or about the existence of objective probability properties of the sort posited by indeterminists—she is making minimal probability statements.

The sort of normativity that depends on the minimal probabilities is *minimal-probability-relative normativity*. It depends on the propositions that (a) render accurate, and that (b) are stood for, by expressions of credences, just as objective normativity depends on the propositions that render accurate and are stood for by expressions of full belief. So when Athena, uncertain among T4, T5, and T6, was asking herself what to do, it is most natural to suppose she was asking herself, "What should I do given that there's some MP that T4 is right, some MP that T5 is right, and some MP that T6 is right?". In other words, she was asking herself what it was MP-relative right to do. This is more psychologically realistic than the supposition that she was asking herself what to do *given her own credences*.

If I'm right that a doxastic state's accuracy condition is simply the proposition that is stood for by what expresses the state, then C2 is intersubstitutable with:

C3: What it's N+1st Order Rational to do depends upon the agent's beliefs or credences whose contents are propositions about a kind of value V, where V depends upon the way the world is according to *expressions* of the beliefs or credences upon which Nth Order Rationality depends (unless N = 1, in which case it depends on the agent's beliefs or credences regarding objective normativity).

And if I'm right that MP-statements are expressions of credences, then C3 is intersubstitutable with:

C4: What it's N+1st order rational to do depends upon the agent's beliefs or credences regarding a kind of value, V, that depends upon the facts (minimal probabilities) that render *accurate* or are mentioned in *expressions* of the beliefs (credences) upon which

Nth order rationality depends (unless $N = 1$, in which case it depends on the agent's beliefs or credences regarding objective normativity).

For agents like Athena, who are uncertain at several levels, it will sometimes be helpful to speak of “orders” not only of rationality, but also of MP-relative normativity. First-order MP-relative normativity will be relative to propositions stood for by expressions of the beliefs/credences to which first-order rationality is relative. *Mutatis mutandis* for second-order MP-relative normativity and second-order rationality, and so on.

2. Systemic Rationality: The First Half of Our Answer

Now that all of the conceptual apparatus is on the table, we're in a position to present our solution to the puzzle about Solomon. Since we've just concluded a long discussion of systemic rationality, let's start with that side of the solution. Here is what our author should say about systemic rationality in Solomon's case: That (1) it is first-order rational, relative to his credences regarding retributivism and its alternatives, to “hedge his bets” and free Norman; and that (2) it is second-order rational, relative to his belief that he MP-relative ought to jail him, to jail him.¹⁸ Nor should it be difficult to see what we might say about Athena.

3. Systemic Rationality: Towards a Complete Answer

But this answer is not entirely satisfying. For we will naturally wonder what's systemically rational, not for some little bit of Solomon's intentional system, but for *Solomon*. In other words, what's globally rational, rather than merely locally rational, to do here?

Let's consider some possible answers:

Answer #1: What's globally systemically rational for Solomon is just whatever is rational relative to Solomon's credences regarding retributivism and the like. This implies that it is rational to free Norman. While it's locally rational, relative to his belief about MP-relative normativity, to jail Norman, what's rational relative to all of the beliefs in question is determined entirely by what's rational relative to his credences regarding objective normativity. The belief about MP-relative normativity is stripped of any influence.

How can we square this answer with the claim that it's indeed second-order rational to jail Norman? Why don't the beliefs to which that kind of rationality is relative play any role in determining global systemic rationality? Perhaps the thought behind this answer is that you shouldn't be able to *change* in any way what's rational for you in circumstances of uncertainty by forming false beliefs about what to do in such circumstances. If we consider in isolation Solomon's belief that he ought to jail Norman, it's fine to say it's rational to jail Norman. But once we consider it as part of a larger intentional system, we may want to take away its “voting power”, as it were.

This thought might gain some support from analogies. Suppose that a thinker believes that P , believes that $P \rightarrow Q$, but thinks the truth of these propositions is compatible with the truth of $\sim Q$.¹⁹ It seems uncontroversibly irrational to believe P , $P \rightarrow Q$, and $\sim Q$ all together, the thinker's crazy belief about the propositions' compatibility be damned. If we say otherwise, then we allow the thinker's belief's about logical compatibility to alter logic's normative role, which seems odd.

The proponent of Answer #1 will say that, in both the theoretical and practical cases, it seems wrong to think that your mistaken beliefs might alter what actually is rational—not for a tiny bit of your intentional system, but for *you*.

But Answer #1 is defective. It claims that there is indeed something that's overall rational for you to do, but accords certain parts of your intentional system no influence whatsoever in determining it, even though those same parts determine what it's locally rational to do. If it is rational relative to some belief to do some action, and this belief is just as much a part of me as any other mental state, how can we justify depriving the belief of any say about what it's overall rational for me to do, whilst according other beliefs such a say?

So let us consider instead:

Answer #2: What's globally rational for Solomon to do is just whatever is rational relative to Solomon's belief about MP-relative normativity. This implies that it is rational to jail Norman. While it is first-order rational, relative to Solomon's credences regarding retributivism and the like, to free Norman, these credences are stripped of any influence over what it is overall most rational for Solomon to do.²⁰

This answer is the complement of Answer #1. Something similar to it is defended in Robert Nozick's *The Nature of Rationality*. Nozick is discussing what it's rational to do in Newcomb's Problem cases when you're uncertain whether Evidential decision theory (which counsels "one-boxing") or Causal decision theory (which counsels "two-boxing") is correct:

"Let CEU(A) be the causally expected utility of act A, the utility that act as it would be computed in accordance with . . . causal decision theory; let EEU(A) be the evidentially expected utility of act A, the utility of that act as it would be computed in accordance with evidential decision theory. Associated with each act will be a decision-value DV, a weighted value of its causally expected utility and its evidentially expected utility, as weighted by that person's confidence in being guided by each of these two kinds of expected utility.

$$DV(A) = W_c \times CEU(A) + W_e \times EEU(A)$$

And the person is to choose an act with maximal decision-value."²¹

On Nozick's view, then, it's irrelevant which of Causal or Evidential decision theory is *actually correct* relative to the agent's beliefs about what's in the boxes. All that matters are the agent's beliefs about the two theories.

Now, Nozick misses out on two ways of making his view more plausible. First, he does not distinguish between local and global rationality. The idea that overall

systemic rationality is determined entirely by one's higher-order views about the rules goes down easier if we can simultaneously grant that rationality *relative to a subset* of one's mental states is determined by the rule of local rationality relative to that subset. Second, Nozick does not distinguish between rationality and MP-relative normativity. A better version of his view would say that Causal and Evidential decision theories can be understood *both* as theories of rationality, and as theories of MP-relative normativity. If understood as theories of rationality, the probabilities they'd take as inputs would be credences. If understood as theories of MP-relative normativity, the probabilities they'd take as inputs would be MP's. Given this distinction, we would say that what's lower-order rational would depend on which of the Causal or Evidential theories of *rationality* is correct, but that what's higher-order rational would depend on the agent's *credences* regarding Causal or Evidential theories of *MP-relative normativity*.

But there seems to be something fishy about Answer #2. For just as Answer #1 says there's something it's overall most systemically rational to do, but makes it entirely insensitive to the agent's beliefs about MP-relative normativity, Answer #2 makes rationality entirely insensitive to the agent's beliefs about objective normativity. On the conception of rationality as appropriate reflection of one's intentional system, then, Answer #2 fails in a way complementary to one of the ways Answer #1 fails.

My suspicion is that this answer seemed so attractive to Nozick either because he was thinking of rationality as perspectival rationality, or else because he was simply running together perspectival rationality and systemic rationality. For it is indeed perspectivally rational for an agent who judges that some choice would have the "maximum decision value" to make that choice, whether or not such a choice is sanctioned by the correct one of Causal and Evidential decision theory.

4. Systemic Rationality: A Positive Proposal

We've seen that Answers #1 and #2 fail as accounts of global systemic rationality. Each says there's something that it's overall rational to do, but consigns some subset of a person's beliefs to "present but not voting" status when it comes to determining this. Each constitutes what seems like an unwarranted privileging of some orders of rationality over others.

Note that it's not hard to come up with a view that avoids such privileging. We might say that the most globally systemically rational action for a person is the action is the most rational *at the most orders* for that person. So if someone has beliefs sufficient for her to be evaluable at five orders of rationality, this view would have it that A is more globally rational than B for her if A is more rational at three orders and B at two, or A at four and B at one, or A at two and B at one with two "ties", and so on. Or we might say that the most globally systemically rational action for a person is whichever has the highest *sum of degrees of rationality of all orders*. (To get A's degree of rationality at an order N, we would normalize that order—that is, assign the most Nth order rational action a value of 1, and the least Nth order rational action a value of zero—and then assign to A

a value between zero and 1 inclusive, in accordance with the value function for that order.)

The difficult task in this section, then, is not simply to come up with a non-privileging view of global rationality. It takes only a little imagination to do that. Rather, the difficult task is to select among the myriad non-privileging views on grounds that go beyond, e.g., quasi-aesthetic intuitions of theoretical elegance. I would recommend a broadly pragmatist strategy for choosing among candidate theories: First, we determine what we need a theory of global systemic rationality *for*—i.e. what roles such a theory is required to play (or not required to play) in thought and action. Second, we select among candidate theories according to whether and how well they play these roles.

The discussion here will be, of necessity, preliminary. For there are many roles that one might plausibly assign to a theory of rationality, and it's a vexed question which theories best enable rationality to fulfill its proper roles. Suppose, for instance, that we wanted an account of global systemic rationality that enabled rationality to play its proper role in a theory of *morality*. For illustration's sake, suppose this role is as follows: we wrong someone if we treat him in a way to which he could not globally systemically rationally consent. To select a theory on this basis, we'd first have to show that global systemic rationality indeed plays this role in morality. Then we'd have to show that the completed moral view we get by filling in the "global systemic rationality" box one way is more satisfactory than the completed moral view we'd get by filling it in any other. A task like this is too big for one section of a paper.

What I'd like to do, though, is to convince you of the *plausibility* of a non-privileging view that some may be inclined to reject out of hand. That is the view that there is *nothing it's most systemically rational* for Solomon to do. It is neither more rational for him to free Norman nor more rational for him to jail Norman. (Whether we want to say that it is rationally *permitted* to do *both* or to do *neither* is something I don't wish to take a stand on here.) This is because one cannot compare differences in degrees of rationality between orders. If one action or thought is more first-order rational for someone, and another is more second-order rational for that person, then neither is more globally systemically rational.

The idea of incomparability is a familiar notion from other arenas. It used to be a platitude in economics that differences in well-being were incomparable between people. So if a policy makes some people better off and others worse off, it was thought to be the case that it neither increases nor decreases aggregate well-being. And of course incomparability is a familiar notion from moral theory. We might say that moral reasons and self-interested reasons are incomparable with respect to overall reason strength, or that goodness as a painter and goodness as an engraver are incomparable with respect to goodness as an artist.²²

The major argument against this position is a pragmatic one: Rational orders simply *have* to be comparable so that there is a single answer to the question, "What is it (systemically) rational to do?"; otherwise certain agents would be unable to engage in guided action. I've heard this view in many conversations, and it is

implicit in the claim the the author “wouldn’t exactly be offering helpful advice” by advancing the third view in the introduction. Now, some philosophers would be undisturbed by this charge, for they think that, as a general matter, believed incomparability poses no threat to guided thought or action.²³ If that were the case, then of course belief in the particular sort of incomparability I’m alleging would pose no such threat. But I don’t want to commit myself to such a strong thesis. Instead, I will advance a more limited claim—that when types of value are related to one another in the specific way that orders of rationality are related to one another, believed incomparability among them poses no threat to guided thought or action. So absent some *other* point to comparing different orders of rationality with respect to global rationality, we have no reason to regard them as so comparable.²⁴

The first-pass argument that we don’t *have* to compare orders of rationality is very simple. As we learned in the discussion of orders of rationality, one needn’t think in terms of rationality to guide one’s attempts at doing what is objectively valuable. One can rely on a combination of thoughts about objective norms and thoughts about MP-relative norms. But if we don’t need to use norms of rationality at all in deciding what to do, then *a fortiori* there’s no practical point in comparing different orders of rationality.

But I can see at least two objections to this first-pass argument. First: “That the orders of rationality are incomparable implies that the sorts of normativity that are the objects of the beliefs upon which these orders depend are incomparable. These sorts of normativity are objective normativity and MP-relative normativity. And this latter incomparability really does pose a threat to guided choice. So we should reject the view that the orders of rationality are incomparable.” Second: “The first-pass argument just given relies on the earlier claim that we needn’t ever guide our behavior by rational norms. But I actually didn’t accept that earlier claim, so I am unwilling to believe that the orders of rationality are incomparable absent some further argument.”

Let me consider these objections in turn.

The first objection is that, since propositions about objective and MP-relative normativity are the *contents* of the beliefs upon which higher-order-rationality depends, then if these sorts of normativity are comparable, then so must the orders of rationality be. We must posit comparability among objective and MP-relative normativity to engage in rational choice under uncertainty, so we must posit comparability among the orders of rationality, too.

An analogy may help. Let us consider different *types* of *objective* reasons. Suppose for argument’s sake that objective moral reasons and objective prudential reasons are comparable with respect to overall objective reason strength. Given this supposition, it would be bizarre to then deny that the sort of rationality that depends on *my beliefs* about objective moral reasons and the sort of rationality that depends on *my beliefs* about objective prudential reasons are comparable with respect to overall rationality. So similarly, incomparability among the orders of rationality that depend upon beliefs about objective and MP-relative normativity requires incomparability among their objects.

It's not obvious that this argument is correct, but it is plausible enough, so let us assume it works. My response to it is simply that we needn't believe in comparability among objective and MP-relative normativity to engage in rational choice-making, and so we ought to regard these as mutually incomparable, just like the orders of rationality.

For we'd only need comparability among objective and MP-relative normativity for this purpose if an agent could ever coherently face what she sees as a *practical conflict* among the types. However, this is impossible. For each order of MP-relative normativity is "practically transparent" to the orders below it, and all of these are transparent to objective normativity. First consider the case of objective normativity vs. first-order MP-relative normativity. There are two possibilities:

- 1) I have a full belief or certainty that I (objectively) ought to do A.
- 2) I am uncertain about what I (objectively) ought to do.

If 1), then I will probably just do A. I can guide my doing it with this full belief or certainty. I suppose if I were really bored, I might ask myself what I objectively ought to do given that I objectively ought to do A. But of course "A" is the only reasonable answer. If 2), then I will be unable to guide my behavior by this uncertainty.²⁵ I will either have to perform an unguided action or else form an attitude of certainty about MP-relative normativity and guide my behavior by *that*. So there is no conflict here, because only the belief about MP-relative normativity is capable of doing the guiding.

Either way, I as an agent can never face what I regard as a practical conflict between what objective normativity counsels me to do, and what MP-relative normativity counsels me to do. So no need to compare the two.

I've sometimes heard in conversation that we face a conflict among objective and probability-relative normativity in so-called "Jackson cases". In such cases, the agent believes that either doing A or doing B is objectively right (she doesn't know which), but that doing C is probability-relative right. The alleged conflict, then, is that different actions are right in different senses of "right". The agent will see objective rightness and probability-relative rightness as pushing in two different directions.

I suppose we can call this a "conflict" if you wish, but I'd want to deny that it is a genuinely *practical* one. For a conflict to be practical, the agent would have to conceive of more than one action under a description that is action-guiding in that choice situation. But this is not so in Jackson cases. The agent would think of more than one action under an action-guiding description if she thought of either A or B under the description "right", but she does not. He thinks of each under the description "may be right". But thinking an action may be right is not generally sufficient to guide one's doing that action. One and only one action, C, is thought of under an action-guiding description.

That Jackson cases are not truly practical conflicts is, I would think, common ground among those who discuss such cases. For it seems to be generally accepted that the responsible agent will simply do C without a second thought; what

philosophers debate about is whether we can explain why this is so by appeal to objective normative notions only,²⁶ or whether we must appeal to subjective normative notions as well.²⁷

Second-order MP-relative normativity bears the same “practical transparency” relationship to first-order MP-relative normativity that the latter bears to objective normativity. To see this, consider two more complementary possibilities:

- 1) I have a full belief or certainty that I (first-order MP-relative) ought to do A.
- 2) I am uncertain about what I (first-order MP-relative) ought to do.

If 1), then I may simply do A. So there is no practical conflict here. If 2), then I cannot guide my behavior by first-order MP-relative normativity at all, but only by a second-order MP-relative norm. So there is no practical conflict there.

In retrospect, none of this should be surprising. There have been many views offered about the relationship between objective and so-called “subjective” value (MP-relative normativity is a type of the latter), but to my knowledge nobody has ever thought to posit some third, “neutral” kind of value with respect to which the two can be compared for the purposes of guiding action. I suspect the reason for this is the tacit recognition that there is simply no practical need to do so. Similarly, there is no need to posit some overall kind of MP-relative value with respect to which the different orders of it can be compared.

The second objection to my incomparabilist thesis can be dispensed with quickly. If you think that we must guide our behavior under uncertainty by norms of rationality, then you are assigning to such norms the role that I have been assigning to MP-relative norms. But in that case, we may simply replace “MP-relative” with “rational” in the practical transparency argument just given, and get an equally good argument against the need to compare objective normativity and the various orders of rational normativity.

Now, again, the discussion in this section has been preliminary. I’ve focused on only one purpose for comparing orders of rationality. But I have not ruled out that there may be other purposes for doing so. And if there are, then we have reason to regard these orders as comparable. I would contend, however, that the burden is squarely on the “comparabilist” about global systemic rationality to come up with a role that rationality could play only if there were always single answer to the question, “What is it globally systemically rational to do?” As we’ve seen, rationalizing explanations don’t require us to compare orders, for such explanations implicate *perspectival* rationality. And as was mentioned in an earlier footnote, the kind of rationality at issue in ordinary conceptual role semantics is *local*, not global. I had bruited the idea, earlier in this section, that morality’s role for rationality might require the comparability of orders. But of course, many moral theories assign no role to rationality—think of utilitarianism—and those that do might assign that role to perspectival rationality, or first-order systemic rationality, or highest-order systemic rationality, rather than global systemic rationality. In light of these facts as well as the failure of the “guidance” rationale for comparability, it is clear that there is no pat answer to the question of why orders *must* be comparable.

5. Perspectival Rationality

Perspectival rationality is less evenhanded than systemic rationality, for it does not aim to take into account all of the states in an intentional system—only the conscious ones that constitute the agent’s perspective. Given Solomon’s perspective, it would be perspectivally rational for him to jail Norman. That’s because jailing Norman not only appropriately reflects the mental state that constitutes this perspective, but also lacks a phenomenology of unguidedness. The phenomenology of Solomon’s jailing Norman would be as of a seamless transition from his (MP-relative) “ought”-judgment to his announcing “I sentence you, Norman . . .”. But the phenomenology of freeing Norman would be very different. It would seem to be an action in contravention of his perspective on what to do. Derivatively, it would be easy to give a reasons-explanation of Solomon’s jailing of Norman. We’d say, “He jailed Norman because that’s what he thought he ought to do in light of his credence in retributivism.” By contrast, no mental state we’ve imputed to Solomon could serve as the explanandum in an rationalizing explanation of his freeing Norman.

It is with respect to perspectival rationality that Nozick was more-or-less right about what it’s rational to think or do. If I’m sure that Causal decision theory is right, it wouldn’t make sense from my perspective for me to choose one box rather than two, and others couldn’t make sense of me if I did.

It is also with respect to perspectival rationality that the “too subjective” answer from the introduction is correct. But the problems we raised for that answer dissipate once we see how it fits within the broader framework I’ve sketched. The first concern was that a view on which whatever makes sense from your perspective is rational cannot make enough room for errors in thought. We imagined an overly permissive Frege saying that the laws of logic do not apply to one who doubts them. But now we see that the perspectival conception of rationality is only *half* of the story, and that one may be systemically irrational—severely so, even—*notwithstanding* one’s perspectival rationality. The second concern was that on a perspectival conception of rationality, we could not make sense of someone deliberating about what to do when he is uncertain. For whatever belief he arrived at would constitute his perspective about what to do, and hence verify itself. But our battery of distinctions allows us make perfectly good sense of this sort of deliberation. When an agent asks himself, “What ought I to do?” under conditions of uncertainty, the “ought” he is employing is most certainly not that of perspectival rationality. If it were, then the answer would be *nothing*, since no action seems *yet* to make sense from his perspective. That’s what prompts him to ask the question. Indeed, as the “orders” discussion showed, he is likely not employing a rational “ought” at all. Rather, he is in typical cases asking what he *MP-relative ought* to do. But because the way he answers this question determines only what is perspectivally rational to do, not what he MP-relative ought to do given the MP’s mentioned in the expressions of the aforementioned credences, there is no threat of a self-verifying answer, and hence no farcical character to his deliberation.

5.1 *How I've Made Things Easier For Myself (by focusing on Solomon), and What Would Happen if I Made Them Harder (by focusing on Athena)*

Solomon was uncertain about objective normativity, but then certain about MP-relative normativity. Because of this certainty, there was a prospective act of his—jailing Norman—of which he and others can make perfect sense. Jailing Norman was perspectively rational for him.

But Athena's uncertainty was more thoroughgoing. For she was uncertain about every order of MP-relativity about which she had any attitude whatsoever. There seems to be no problem giving an account of systemic rationality in her case. We just say what's locally rational relative to each level of belief, and then say that all of these orders of rationality are incomparable with one another. But if nothing Athena can do would seem natural—if all of her possible actions would seem like leaps of faith—what should we say is perspectively rational for her?

One possible answer is: nothing, and that's all there is to it.

But we might pursue a more subtle answer. My view is that my act is rationalized in the perspectival sense by a set of mental states just in case it is intentionally explained by those mental states. Had Athena fully believed in T1, which, let's say, ranks action A above action B, that belief would have perfectly intentionally explained her doing A. But in fact, she is uncertain among T1, T2, T3. Let us suppose that T1 ranks her prospective actions (from best to worst) as: A, B, C; T2 ranks them: C, B, A; and T3 ranks them: B, C, A. This state of uncertainty does not, then, intentionally explain the performance of either of A, B, or C rather than the others. That is, it would not *contrastively* intentionally explain Athena's doing any of them. But suppose that all three of these moral views ranked D lower than A, B, and C. Then this state of uncertainty might contrastively intentionally explain my doing A *rather than* D or B *rather than* D or C *rather than* D. So we might say that A, B, and C are each *partially* intentionally explained by a state of uncertainty, and hence that each is *more* perspectively rational than D.

I hope to eventually show that the further we “step back” from states of uncertainty—that is, the more orders of MP-relativity we consider—the fewer prospective actions will be left in the disjunctive set amenable to intentional explanation. Correspondingly, the more we step back, the more actions will be “hived off”, or placed in the contrast set. Suppose, for example, that I can do any of mutually exclusive actions A...Z. Perhaps my uncertainty regarding objective normativity will intentionally explain my doing any of A...R rather than S...Z. Then S...Z have been hived off. Now suppose that I consider what to do relative to the MP's expressed in that uncertainty, and am uncertain as to the answer. Perhaps this new uncertainty will intentionally explain my doing any of A...G rather than H...Z. Not only S...Z, but also H...R, have now been hived off. My hope is to show that, as a general matter, potential actions will be hived off with each stepping-back, and that previously hived off actions will never reappear among the set amenable to intentional explanation. I have reasons to suspect that such a showing is possible, but I'll have to make good on that suspicion elsewhere.

6. Other Applications

We might be suspicious if the framework we've developed were just a "one-off" solution to a single, isolated problem. It might make us think that it didn't really address what was deep and important about that original problem, but rather traded on a superficial feature of the problem's construction. For this reason, and frankly, because I think it'd be fun, I'd like to conclude by seeing how a framework something like the one I've developed might be cross-applied to some "hot" topics in epistemology.

6.1 Conciliationism about Peer Disagreement

Suppose my initial credence in P is 1.0, but I believe that Hoon is my peer and has a credence of 0 in P. And suppose further that I accept a version of "conciliationism" about peer disagreement according to which it would be rational for me to adopt a credence in P halfway between 1.0 and 0—that is, .5.²⁸ But of course, whether this theory itself is true is just one more possible locus of disagreement. So let us suppose that I further believe that Imogen is my peer, and has a credence of 0 in this version of conciliationism. Then by the theory's own lights, I should be doubtful of it.²⁹

Now imagine a philosopher who is a defender of this form of conciliationism. What should she say about what it would be rational for me to do in this situation? On one hand, perhaps she should say that it would be rational for me to adopt a credence of .5. After all, this is what we just said her theory implies. The fact that I now doubt this version doesn't make this theory false, any more than my doubt about a scientific theory would make it false. (That should have a familiar ring.) On the other hand, perhaps she should say that I should adopt some other credence. After all, her theory implies that I should now be doubtful *of that very theory*. Rationality is about what makes sense from the subject's point of view and from my point of view this version of conciliationism is quite possibly false. So surely I shouldn't revise my beliefs in accordance with it. (So should that.)

So what should our philosopher say? The worry is that she could not, *qua* defender of her theory, avoid counseling what seems like irrationality. And this, her philosophical opponents would argue, is grounds for rejecting her theory.

I know of two conciliationist responses to this worry in the literature. Adam Elga denies that I should change my beliefs about conciliationism upon my disagreement with Imogen. Rather, a coherent conciliationism, like coherent inductive methods generally, is *self-excepting*. So while conciliationism may require me to revise my beliefs about other matters, my credence in conciliationism should remain 1, come what may.³⁰ David Christensen's view is that believing what is rational according to conciliationism accords with an "epistemic ideal", but that this ideal may conflict with one that counsels respecting one's beliefs/credences regarding conciliationism. It is an open question whether any "Uber-rule" can successfully bring these ideals together.³¹

Without entirely disparaging either of these responses, I humbly suggest an alternative for the conciliationist. She should make the distinction between systemic

and perspectival rationality and recast her theory as follows: It is most locally systemically rational, relative to my initial credence in P and my beliefs about Hoon, to switch to a credence of .5 in P. But, consistently with this, is also most locally systemically rational, relative to the credence in this version of conciliationism I formed after applying this version to my disagreement with Imogen, to adopt some other credence in P. Degrees of these local rationalities are incomparable in the way discussed above, so there is no credence that it is most globally systemically rational to form. But this does not give rise to what I, the thinker, would regard as a true conflict of theoretical reasoning. For I cannot base a revision of my beliefs *solely* upon my initial credence in P and my beliefs about Hoon. Rather, I must employ my beliefs about conciliationism and competing theories of belief revision in the face of disagreement, and so that is what I will do, even if these beliefs are not entirely correct. By contrast, what it is most perspectively rational to believe will depend on the beliefs about theories of disagreement that I form after my encounter with Imogen.

If Christensen's view is the "conflicting ideals" view, we might think of this new view as the "non-conflicting" ideals view. Admittedly, this view is complicated, but appropriately so if the structure of rationality is as fine as I claim it is. We should be leery of answers that simply mush all of the dimensions I've been distinguishing into a single metric of rationality.³²

6.2 Deduction and Doubt

In a recent paper, Joshua Schechter asks us to "[c]onsider a thinker who deduces a conclusion from a justified initial premise through an incredibly long sequence of small competent deductions." "Surely," Schechter claims, "such a thinker should suspect that he has made a mistake somewhere."³³ Let us suppose that, indeed, the thinker suspects that he made a mistake. Assuming he has no other grounds for believing the conclusion, what should we say about the rational status of believing it? On one hand, we might say that this is rational. For it is stipulated that he *actually did* perform the deductions correctly. That he doubts this doesn't make it false ("any more than disbelief in a scientific theory . . ."). On the other hand, we might say that it is not rational. For rationality is about what makes sense from one's perspective. So how can it be rational to adopt a belief based on reasoning that, from one's perspective, is mistaken?

Opinions are divided on this question. Schechter thinks that believing the conclusion is not rational.³⁴ David Hume adopted this stance as well in his *Treatise on Human Nature*, prompting him to declare that all knowledge, even of the *a priori* sort, is "mere probability".³⁵ Timothy Williamson takes a position that, while framed in terms of knowledge, can be translated into the language of rationality so that it conflicts with the Hume/Schechter position: It is rational to believe the conclusion. What's *not* rational is to believe that one is rational in believing the conclusion. To think that the second claim disparages the first is to fall victim to what Williamson calls a "Confusion of Levels".³⁶

I think that both sides have a piece of the truth. I'd want to say that it's locally rational, relative to the initial premise (and perhaps the intermediate steps) to

believe the conclusion. But it is not locally rational, relative to my belief that the conclusion does not actually follow from the premise, to believe the conclusion. (The Williamson-inspired position that rationality does not depend on beliefs about rationality is, for reasons suggested in the “orders” section above, exactly right. My claim is, rather, that local rationality may depend on the “outward-looking” belief that the conclusion *does not follow* from the premise). It is, furthermore, neither more systemically rational to believe the conclusion nor more systemically rational not to believe it. This is unproblematic, because I won’t experience any conflict regarding whether to believe the conclusion. I’ll simply not believe it (although in doing so I’ll be violating a requirement of local rationality). Regarding perspectival rationality, though, it is uniquely rational not to believe the conclusion. This is the sense in which Schechter is correct.

Notes

¹ For their helpful feedback I thank Benj Hellie, Tom Hurka, Zachary Irving, and Karl Schafer, as well as audiences at Toronto, Stockholm, Nottingham, Leeds, and the St. Louis Annual Conference on Reasons and Rationality. This research was supported by the Social Sciences and Humanities Research Council (SSHRC).

² See, e.g., Ted Lockhart, *Moral Uncertainty and Its Consequences* (Oxford UP, 2000); Jacob Ross, “Rejecting Ethical Deflationism,” *Ethics* (2006); Andrew Sepielli, “What to Do When You Don’t Know What to Do,” *Oxford Studies in Metaethics* (2009), and “Moral Uncertainty and the Principle of Equity Among Moral Theories,” forthcoming in *Philosophy and Phenomenological Research*. The question was also among the most hotly debated in the casuist tradition. See, e.g., Gabriel Vasquez, *Commentaria in II-II Summa Theologiae Sancti Thomae Aquinatis* (Alcala, 1598–1615); and St. Alphonsus Liguori, *Theologia Moralis* (2nd ed. 1755) (Reformation Society, 1852). See Albert Jonsen and Stephen Toulmin, *The Abuse of Casuistry* (California UP, 1990) for a summary of this work. Finally, I was pleased but not surprised to see a sophisticated discussion of the issue by David Foster Wallace in “Authority and American Usage,” in *Consider the Lobster* (2006).

³ Gottlob Frege, *Grundgesetze der Arithmetik, Begriffsschriftlich Abgeleitet*, 2 vols. (1893, 1903), partial translation in *The Basic Laws of Arithmetic: Exposition of the System*, trans. M. Furth (California U.P., 1964).

⁴ Not Frege.

⁵ This strand includes John Broome, *Rationality Through Reasoning* (ms); Niko Kolodny, “Why Be Rational,” *Mind* (2005); Nic Southwood, “Vindicating the Normativity of Rationality,” *Ethics* (2008); John Brunero, “On The Scope of Rational Requirements,” *Philosophical Quarterly* (2010); Jonathan Way, “The Symmetry of Rational Requirements,” forthcoming in *Philosophical Studies*. For instance, Kolodny writes: “When we say ‘you ought to’ in the sense of ‘you have reason to’, we usually seem to be saying something about the relation between your *situation* and your *attitudes*. When we say ‘you ought to’ in the sense of ‘it would be irrational of you not to’, we seem to be saying something about the relation between your *attitudes*, viewed in abstraction from the reasons for them.” (2005, p. 455).

⁶ Wittgenstein, *Philosophical Investigations*, 4th edition, Anscombe, Hacker, and Schulte, trans. (Wiley-Blackwell, 2009), section 159.

⁷ For argument on this point, see my “Subjective Normativity and Action Guidance,” *Oxford Studies in Normative Ethics*, Vol. II, ed. Mark Timmons (Oxford University Press, 2013).

⁸ Nor is it difficult to imagine a *revision-of-mind* that is systemically rational but not perspectively rational. Suppose that Thelma has a set of beliefs and credences S_1 , which range over various and sundry propositions. Some formal epistemologist proposes an ultra-complex belief-update rule, R , according to which, if one holds S_1 at T , one should hold S_2 at $T+n$. If this rule is right, then it is systemically rational, relative to S_1 at T , for Thelma to form S_2 at $T+n$. But this is not entirely perspectively rational. For S_2 will not be uniquely intelligible from Thelma’s perspective. To move from S_1 to S_2 in such a way

that her forming S2 is not stab in the dark, she will have to form a further belief *about what to believe* or *what is likely true* given the contents of the states in S1. And if we want to intentionally explain her formation of S2, we cannot simply cite S1. S1 does not give us enough to understand why Thelma formed S2, rather than S3 (which accords with some other fancy-shmancy belief-update rule), or S4 (which accords with another), and so on.

⁹ Especially helpful here are Kolodny's "Why Be Rational" and "State or Process Requirements," *Mind* (2007).

¹⁰ For example, the very interesting holistic view that John Brunero proposes in his "On the Scope of Rational Requirements," makes sense on the "is explained by" interpretation, but not on the "supervenes upon" interpretation.

¹¹ These local rules are also the ones used in non-holistic conceptual role meta-semantic theories. For example, when Ralph Wedgwood suggests that accepting "X is better than Y for purpose P" rationally commits one to having a "preference-with-respect-to-P for X over Y", the notion of rationality employed seems clearly to be local rather than global. See his "Conceptual Role Semantics for Moral Terms," *Philosophical Review* (2000).

¹² Ross, *The Right and the Good* (1930) (Oxford UP, 2002).

¹³ A referee suggested that we could dispense with the need for local rationality by adopting a "wide-scope" view of rationality. On such a view, presumably, we would impose only the following rational requirements on Solomon: that (a) it is *irrational* for him to jail Norman whilst being uncertain as he is regarding retributivism; and that (b) it is *irrational* for him to free Norman whilst being certain as he is that his favored rule for action under uncertainty is correct (and that our imagined author's rule is incorrect). Given only these requirements, it's perfectly possible for Solomon to act without his action being in any way irrational. For he might either (a) jail Norman while *giving up* his credence distribution regarding retributivism, or (b) free Norman while *giving up* his belief regarding what to do under uncertainty.

This is a challenging point. I can do no better than to present a response adapted more-or-less straight from Kolodny's "Why be Rational?": For several reasons, Solomon cannot rationally give up his credence in retributivism based on his current failure to jail Norman. He might be able to rationally give up this credence based on some *other* mental state—e.g. a belief regarding the persuasive force of some anti-retributivist argument. But giving up this credence takes at least some time; Solomon is stuck with it for (literally) now. So Solomon is being irrational (again, locally) by not initiating his jailing of Norman now, and until his credence is given up. *Mutatis mutandis* for his belief about what to do under uncertainty and the action of freeing Norman.

¹⁴ It is crucial to distinguish between this characterization, and one on which the rationality of an action is the measure of the quality of the action *qua* try at doing what would be objectively *right*, or at doing what the agent objectively *ought* to do. "Jackson" and "mineshaft" cases are counterexamples to these other formulations, but not to mine. I explain why in my "Subjective Normativity and Action Guidance". For more on Jackson and mineshaft cases, see Frank Jackson, "Decision Theoretic Consequentialism and the Nearest-and-Dearest Objection," *Ethics* (1991), and Derek Parfit, *On What Matters* (Oxford UP, 2011), respectively.

¹⁵ I say "generally" because this negative appraisal of C1 admits of an exception. On C1's treatment of *first-order* rationality, the rational action at this order depends on the agent's beliefs or credences regarding objective normativity. And it seems almost trivial that the best try for an agent will depend on the same.

¹⁶ For more discussion on this point, see my "Subjective Normativity and Action Guidance".

¹⁷ See my "Subjective Normativity and Action Guidance". For more sophisticated attempts to develop an expressivist semantics for probability statements, as well as epistemic modal statements, see Eric Swanson, *Interactions with Context*, M.I.T. Ph.D. dissertation (2006); Seth Yalcin, "Epistemic Modals," *Mind* (2007) and "Nonfactualism about Epistemic Modality," in Egan and Weatherston, *Epistemic Modals* (2010); Benjamin Schnieder, "Expressivism Concerning Epistemic Modals," *Philosophical Quarterly* (2010); and Sarah Moss, "Epistemology Formalized," forthcoming in *The Philosophical Review*.

¹⁸ Anyone who still has doubts about MPs or MP-relative normativity may simply substitute whatever kind of normativity he or she thinks "V" is in C2 and/or C3 above.

¹⁹ If it initially strains credulity to suppose a thinker might believe these are compatible, then either (a) consider one who believes this for very deep theoretical reasons, or else (b) replace the beliefs in P and $P \rightarrow Q$ with more, or more logically complex, beliefs. Thanks to a referee for help on this point.

²⁰ That is, except insofar as they supply the “raw materials” for probability-relative normativity, in the same way that the world generally provides the “raw materials” for objective normativity.

²¹ Nozick, *The Nature of Rationality* (Princeton UP, 1993), p. 45.

²² For the idea that comparisons of different values make sense only with respect to a “covering value”, see Ruth Chang, “Introduction,” *Incommensurability, Incomparability, and Practical Reason* (Harvard UP, 1997), and “Putting Together Morality and Well-Being,” in P. Baumann and M. Betzler, *Practical Conflict: New Philosophical Essays* (Cambridge UP, 2004).

²³ See, e.g., Michael Stocker, *Plural and Conflicting Values* (Oxford UP, 1990).

²⁴ I owe this “pragmatist” way of thinking about incomparability to Elizabeth Anderson, *Value in Ethics and Economics*, (Harvard UP, 1995).

²⁵ See my “Subjective Normativity and Action Guidance” for an extended discussion of why this is.

²⁶ See, e.g., Peter Railton, “Alienation, Consequentialism, and the Demands of Morality,” *Philosophy and Public Affairs* (1984), as discussed in Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”.

²⁷ See, e.g., Frank Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.”

²⁸ Note that this is conciliationism about rationality, specifically. This should not be taken to imply conciliationism with regard to other epistemic notions. For example, it does not imply, and I do not endorse, the view that the *evidence supports* a credence of .5 in the case described. As I argue in my “How to Think about Evidence, Rationality, and Disagreement” (ms), conciliationism is true only with regard to rationality and certain notions—there’s a more “external” one I call “reasonableness”—that are defined in terms of rationality.

²⁹ Brian Weatherson presents situations like this as a problem for conciliationism in “Disagreeing about Disagreement” (ms).

³⁰ Elga, “How to Disagree about How to Disagree,” in R. Feldman and T. Warfield, *Disagreement* (Oxford UP, 2010).

³¹ Christensen, “Epistemic Modesty Defended” (ms).

³² This problem, as I’ve understood it, should not be confused with this related problem: I came to adopt a new credence in conciliationism by applying conciliationism to the disagreement between me and Imogen. Given this new credence, shouldn’t I go back and revise my response to *this very disagreement*? After all, it was the result of applying a rule of which I’m now doubtful. This of course will give me a *newer* credence in conciliationism, which should, again, prompt me to go back and revise my response *again* to this disagreement. Where does this process end? Do I keep revising my credence in conciliationism and then returning to the disagreement about conciliationism armed with this new credence? Is there a stable equilibrium credence? Several stable equilibrium credences? I think the framework I’ve developed here could be useful in addressing this problem, but I don’t have a worked-out solution to it yet. For discussion of this “cycling” problem, see Matthew Weiner’s blog post at: <http://mattweiner.net/blog/archives/000780.html>.

³³ Schechter, “Rational Self-Doubt and the Failure of Closure,” *Philosophical Studies* (2011). See also Maria Lasonen Aarnio, “Single-Premise Deduction and the Failure of Closure,” *Philosophical Studies* (2008), who argues that single-premise and multi-premise deductions ought to be treated alike for these purposes.

³⁴ Schechter, “Rational Self-Doubt”.

³⁵ Hume, *A Treatise on Human Nature*, I.IV.I.

³⁶ Williamson, “Improbable Knowledge” (ms). See also William Alston, “Level Confusions in Epistemology,” *Midwest Studies in Philosophy* (1980).