INSTRUMENTAL JUSTIFICATION OF EPISTEMIC COHERENCE REQUIREMENTS

John Phillips

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Philosophy in the College of Arts and Sciences.

Chapel Hill
2015

Approved by:

Matthew Kotzen

Ram Neta

L. A. Paul

ABSTRACT

JOHN PHILLIPS: Instrumental Justification of Epistemic Coherence Requirements.
(Under the direction of Matthew Kotzen)

Why think that our doxastic states should obey norms of formal coherence? One sort of answer to this question attempts to justify coherence in instrumental terms, arguing that only by being coherent can we achieve some more fundamental goal, such as true belief. Here I consider apparent counterexamples to this instrumental strategy: cases in which being *incoherent* seems like the best way of achieving epistemic goals. Although the cases differ considerably in their surface structure, I provide a common diagnosis, claiming that the cases implicitly depend on contentious assumptions about the range of possible states that count as relevant alternatives to be considered in pursuing our epistemic goals. I suggest that, by substituting a range of such states that is independently better motivated, we may preserve the viability of the instrumental strategy.

ACKNOWLEDGMENTS

I would like to thank Matt Kotzen, Ram Neta, and Laurie Paul for invaluable discussion in the course writing this thesis, and Matt in particular for his generosity in devoting a great deal of his time to the project. I'd also like to thank my parents for their support, the members of my cohort at UNC Philosophy for being a constant source of commiseration, and Courtney Monahan for her encouragement and her incredible patience in putting up with me throughout.

TABLE OF CONTENTS

INTRODUCTION

Consider the following two claims:

TB: One ought to believe as many truths, and as few falsehoods, as possible.

NC: For any proposition p, one ought not simultaneously believe p and $\neg p$.

Each of these claims expresses a candidate epistemic norm, a proposal about how an agent's doxastic attitudes must be in so far as she is to count as epistemically rational. Perhaps neither is obviously correct; TB in particular is clearly in need of greater specificity (how, for instance, is one to trade off between the dual demands of believing truths and avoiding believing falsehoods?). But each, I think, has at least a certain ring of plausibility to it, a sense that if we haven't yet stated things in quite the proper way, we're at least on the right path.

In certain cases, TB and NC conflict; in order to satisfy one, you must violate the other. Simply imagine any case in which your holding a pair of contradictory beliefs results in your acquiring (only) a great many new true beliefs, or losing (only) a great many false ones. Some such cases are simple but whimsical: a mad scientist offers to dose you with a drug that will greatly increase your deductive powers, on the condition that his brain scanning machine detects in you simultaneous beliefs that seven is a prime number, and that it is not. Other cases are more realistic but also require more detailed exposition; I will present some examples later. In such cases, there are a number of interesting questions that can be asked: is it *possible* for you to come to believe a contradiction? Is it *prudent* to do so? Is it *morally permissible*? But we can also ask distinctively epistemic questions: is it *epistemically rational* for you to hold the contradictory beliefs, violating NC? Is it rational not to, thereby violating TB? Both? Neither?

At this point, you might reasonably question the import of a conflict between two candidate norms. After all, they are (for all that has been said, anyway) only *candidate* norms; the fact that they conflict might simply be evidence that at least one of them is incorrect, and that their intuitive plausibility is at best a reason to seek related alternatives that do not conflict. And even if we are convinced that NC and TB are genuine epistemic norms — well, norms sometimes conflict. We could still try to figure out an appropriate response to such situations, or simply hope that we ourselves are lucky enough to avoid such perplexing circumstances.

I think, though, that such a response does not do justice to the importance of this conflict. TB and NC were not picked at random; instead, they are representative of two broader categories of candidate norms, what I will refer to as 'goal norms' and 'coherence requirements' respectively. I will discuss these categories at length in Section I, below. The conflict sketched above is not specific to TB and NC, but arises for a great many combinations of a goal norm and a coherence requirement. Such conflict poses a challenge to an otherwise promising (though certainly not the only) strategy both for accounting for the apparent normative force of coherence requirements, and for figuring out which coherence requirements in fact hold. This strategy, which I will term the 'Instrumental Thesis,' suggests that coherence requirements hold in virtue of their status as a means to the end of conforming to goal norms. This thesis, if true, can be used to explain, of a given coherence requirement that we are convinced holds, *why* we ought to follow it: in order to achieve the end represented by some goal norm. Or the thesis can be used, given a goal norm, to figure out what coherence requirements must hold— we must simply figure out which coherence requirements serve as effective means to the end of conforming to the goal norm.

The challenge, in short, is this: whatever epistemic coherence requirements in fact hold, they seemingly hold *unconditionally*.[1] Intuitively, for instance, there is *some* requirement that you are violating in holding contradictory beliefs, regardless of the circumstances. In particular, coherence requirements seem to hold whether or not they are effective means, in a given instance, of satisfying a goal norm. If it is true that for any goal norm and any coherence requirement there are cases where they conflict, then there are cases where the coherence requirement is not an effective means of achieving the goal norm. If the coherence requirement derived its normative force from its status as an effective means, then we would expect it not to hold in these cases of conflict. But then it is would be false that the coherence requirement holds unconditionally. This might, of course, turn out to be the right answer; intuitions to the contrary might be in error.

The main project of this paper will be to consider what may be said in favor of coherence requirements in light of this challenge. What resources might we draw on to resist abandoning unconditional coherence requirements altogether? I will not defend a specific positive argument for a particular coherence requirement, but will instead try to show where in logical space such arguments may be found. While I give some consideration to the option of giving up the Instrumental Thesis in favor of alternative strategies of justification, my main focus will be on means by which the Thesis itself may be preserved: in particular, on the characteristics of goal norms that do not under any circumstances come into conflict with the coherence requirements that they might be used to justify.

I will proceed as follows. In Section I, I will clarify the categories of 'goal norms' and 'coherence requirements.' In Section II, I will motivate the Instrumental Thesis, and show how it has been used to argue for epistemic coherence requirements. In Section III, I will present two cases of apparent conflict between coherence requirements and goal norms, of the sort that pose a challenge for the Instrumental Thesis. Section IV will provide a

---

[1] Indeed, given the definition that I will offer below, a norm must hold unconditionally to count as a coherence requirement at all, though of course a given coherence requirement may be closely related to a conditional norm that otherwise restricts one's doxastic states in similar manner.

common diagnosis for these two cases, arguing that in each case the root of the conflict lies in the interpretation of modal terms used in the goal norm, and offer a general account of the conditions under which conflicts between coherence requirements and goal norms can arise. Section V will consider the prospects for using the Instrumental Thesis to justify coherence requirements in light of this general challenge. (Sections IV and V together constitute the argumentative heart of the paper.) Finally, Section VI will consider the merits of two alternative responses to the challenge: either denying the existence of unconditional coherence requirements, or rejecting the Instrumental Thesis as an explanation of their force.

1 GOAL NORMS AND COHERENCE REQUIREMENTS

TB is one way of spelling out the adage that beliefs 'aim at the truth,' where that aiming is taken to be a normative rather than a purely descriptive matter. TB picks out (when relativized to an agent's circumstances) a certain ideal set of beliefs: the set including as many true beliefs and as few false ones as it is possible to hold under the circumstances.[2] Of course the norm does not pick out this belief set under a sufficiently informative description to settle the question, from the perspective an actual agent, of which propositions she should believe; it does not tell her which are true and which false. Thus, knowing that she is subject to TB and intending to satisfy it does not ensure her success. The norm does, however, establish the *correct* answers, and provide a standard against which her beliefs can be objectively evaluated.

We can imagine altering TB in several ways while remaining consistent with its general spirit. The norm might be supplemented so as to specify various ways in which the sub-goal of maximizing true beliefs should be traded off against that of avoiding false beliefs, in non-ideal cases. For instance, we might specify that either holding true beliefs or avoiding error takes lexical priority over the other, or that the acceptable expected 'cost' of acquiring a true belief is some particular number of false beliefs, or that acquiring true beliefs about certain particularly important topics is worth taking on false beliefs about less significant issues.

TB might instead be modified so as to pick out a different set of ideal beliefs, perhaps one not specified in terms of truth and falsehood; say, the set of beliefs best supported by

---

[2]How we are to understand just which belief sets count as 'possible to hold under the circumstances' will turn out to be a pivotal question. As this statement of TB is open to multiple readings, it can be more precisely understood as representing a family of closely related goal norms. At this stage, however, my aim is to demonstrate the structure of goal norms, not to identify the best version of TB.

one's current body of evidence. (Note that asserting this norm is quite different from simply pointing out that a rational agent attempting to satisfy TB will make use of her evidence to do so, and that her ability to come to true beliefs will depend on what evidence she has available. The evidential norm in question states that a rational agent *ought* to believe what is supported by her evidence, even if what is supported by her evidence is also false; and she may fail to perfectly satisfy it if she reaches conclusions that are not in fact best supported by her evidence, even if they are in fact true.)

Or again we might suggest norms analogous to TB, but applying to doxastic attitudes other than full belief. A prominent example is the accuracy norm for credences (discussed in greater detail in Section II), but perhaps there are also plausible analogs for comparative confidences, suppositions, and so forth. A goal norm might consist of a conjunction of such claims, for instance the conjunction of TB for beliefs and an accuracy norm for credences.

All of the norms resulting from these alterations (applied in any combination) would be what I'll refer to as 'goal norms.' A goal norm in the sense used here is distinguished by the fact that it specifies a certain ideal state, or set thereof, that an agent ought, as a matter of epistemic rationality, to strive to come as close to occupying (one of) as possible. Goal norms are expressible as conditionals (or conjunctions thereof) with the 'ought' taking narrow scope. For instance, TB can be expressed as follows:

> For any belief set B, if B is the belief set containing as many true beliefs and as few false beliefs as it is possible for one to believe, then one ought to hold B.

(Again, TB must be specified more precisely in order to establish a fact of the matter about which belief set satisfies the antecedent of this conditional).

There's a point worth briefly dwelling on here for the sake of clarity. A goal norm may stipulate its ideal state solely in terms of the intrinsic features of an agent's doxastic attitudes. However, many examples, including TB, don't work like that. The contents of the ideal belief set picked out by TB depends on which propositions are true and which are

6

false, that is, on facts about the world; whether one has satisfied TB depends *both* on one's doxastic attitudes, *and* on those worldly facts. Thus, in contrast to the coherence norms addressed below, many goal norms are extrinsic in the sense that no doxastic state that an agent can occupy is *in itself* sufficient to guarantee that they conform to a given goal norm; rather, that doxastic state must bear an appropriate relation to the relevant facts. Satisfying these norms looks different depending on the features of the world that one is in.

A natural way of understanding goal norms is in terms of *epistemic value*. For instance, we might understand TB as specifying sources of positive (true beliefs) and negative (false beliefs) epistemic value, and requiring that rational agents adopt the doxastic state with as much total epistemic value as possible. Other goal norms might be spelled out not in terms of maximizing value, but rather of achieving sufficient value, or of avoiding value dominance. In many contexts, such talk of epistemic value will be equivalent to talk of goal norms as presented in the previous paragraph. However, I do not mean to assume that the value involved in a genuine goal norm must necessarily be the only sort of epistemic value, or that satisfying goal norms must be the only source of epistemic value; the formulation of goal norms in terms of ideal states is meant to avoid this suggestion.

NC is a very different sort of norm from TB. It is an example of what I will refer to as 'coherence requirements'. Coherence requirements make claims about the *patterns* that should or should not hold among an agent's doxastic attitudes, rather than about whether any particular attitudes should or should not be held. One feature of coherence requirements, then, is that that they can be stated as conditionals (or conjunctions thereof) with the 'ought' taking a wide scope (albeit sometimes at the expense of standard English grammar). For instance, NC was stated at the outset of the paper as follows:

> For any proposition p, one ought not simultaneously believe p and $\neg p$.

Here is the equivalent in wide-scope conditional form:

> For any proposition p and time t, one ought (if one believes p at t, then one does not believe $\neg p$ at t), and one ought (if one believes $\neg p$ at t, then one does not believe p at t).

This coherence requirement is distinct from the narrow-scope claim NC*:

NC*: For any proposition p and time t, if one believes p at t, then one ought not believe $\neg p$ at t, and if one believes $\neg p$ at t, then one ought not believe p at t.

One important difference is that the consequent of the conditional in NC* can be detached via modus ponens; if one in fact believes p at t, then it follows from NC* that one ought not believe $\neg p$ at t, simpliciter (even though one is perhaps mistaken in believing p, and $\neg p$ is true, better supported by one's evidence, etc.). In contrast, NC offers no guidance at all about whether one ought to believe p or $\neg p$. It simply says not to believe them both at the same time.

Not all wide-scope epistemic norms are coherence requirements. Coherence requirements speak in particular to patterns among one's doxastic attitudes alone, in the sense that the truth of both the antecedent and the consequent of each conditional in the conditional form of a coherence requirement depend solely on what doxastic attitudes one has. There are (at least possible) wide-scope epistemic norms that speak instead to the relationship between one's doxastic attitudes and other facts about the world; these involve conditionals where only the consequent depends on one's doxastic state. For instance, the following might be an epistemic norm: one ought (if one sees a squirrel, then one believes there is a squirrel present). This norm is wide-scope, and is defensible, but is not a coherence requirement: what doxastic attitudes one should have, in order to satisfy it, depend on whether or not one actually sees a squirrel.[3] Of course, the squirrel norm is not a goal norm either; goal norms and coherence requirements are not meant to exhaust the space of logically possible epistemic norms.

For purposes of this paper, I will be restricting my consideration to synchronic coherence requirements. Besides NC, examples of widely defended coherence requirements include

---

[3]I am assuming a sense of 'sees' here that does not itself count as a doxastic attitude.

Probabilism: One ought not assign a credence function that violates the axioms of probability.

and

Deductive Consistency: For any propositions p, q and r, if r is a logical consequence of p and $\neg r$ is a logical consequence of q, then one ought not simultaneously believe p and q.

Of course, many other coherence requirements are possible. I will focus my discussion, in particular, on NC and Probabilism, but the main points generalize at least over synchronic epistemic coherence requirements.

2   THE INSTRUMENTAL THESIS

The Instrumental Thesis is a claim about the relationship between epistemic coherence requirements and epistemic goal norms. According to the Thesis, any genuine coherence requirement holds in virtue of the fact that satisfying the coherence requirement is a means of bringing about the end of satisfying some genuine goal norm. The force of the coherence requirement is therefore derivative of the force of the goal norm: you ought to satisfy the former *in order* to satisfy the latter. This leaves open the question of whether the goal norm is fundamental in an absolute sense, or whether it in turn holds in virtue of its relation to some yet more fundamental norm or source of value; nonetheless I will speak of the *more* fundamental norm in a means/end relationship as 'fundamental.'

I noted at the beginning of Section I that TB is one way of making sense of the claim that beliefs 'aim at the truth.' In general, goal norms that specify the ideal epistemic state as depending both on one's doxastic attitudes and on the truth values of the propositions over which one is opinionated (alternatively, norms that take truth to constitute epistemic value) have an easy time explaining the sense that, as believers, we are in the business of trying to get at the truth. Of course, there can be considerations in favor of holding some belief that have nothing to do with its truth. A belief may be prudentially, or morally, or perhaps even aesthetically valuable. But one might think that *epistemically speaking*, facts about what beliefs we ought to have trace back in some manner or other to facts about which beliefs are true and which are false. Claims roughly along these lines are common in the literature.[4] William Alston, for instance, writes: "Epistemic evaluation is undertaken from what we might call 'the epistemic point of view.' That point of view is defined by the

---

[4]See for instance David [2001], Alston [1985], Bonjour [1985].

aim of maximizing truth and minimizing falsity in a large body of beliefs .... For a belief to be justified is for it, somehow, to be awarded high marks relative to that aim."[5] While Alston puts the point in terms of justification, in the language of norms we can read him as claiming that any norm governing beliefs that has nothing whatsoever to do with promoting *true* belief should not be counted as an *epistemic* norm at all.

My purpose here is not to legislate the use of the term 'epistemic.' I will not argue that applying that description to a norm that in no way involves promoting true belief, or even promoting any particular relation between one's doxastic state and the world, must be mistaken. Still, whether or not what Alston gestures at is appropriately considered *the* epistemic point of view, it is certainly *an* epistemic point of view, and a familiar, compelling one at that. If we embrace a goal norm like TB, then we are firmly situated within this familiar epistemic point of view. If we adopt a goal norm that differs *too* radically from TB, then whether or not the point of view we thereby take still counts as an epistemic one, we must consider whether the familiar point of view is one that we are willing to abandon.

What then to say about coherence requirements? Let us assume for the moment that if coherence requirements like NC and Probabilism hold at all, they are *epistemic* norms. If that's so, and if we accept that epistemic norms must somehow be in the business of promoting true belief, then we apparently are faced with a puzzle: *how* are NC and Probabilism in the business of promoting the truth? Recall, neither one makes any mention of truth at all. Indeed, I stipulated that coherence requirements, as I am using the term, are satisfied (or not) based solely on facts about what doxastic attitudes one has. Apart from perhaps a few tricky cases of beliefs about one's own beliefs, this stipulation rules out the possibility that whether a coherence requirement is satisfied depends directly on the truth of one's beliefs. Are we then forced to give up on the idea of genuine epistemic coherence requirements altogether?

---

[5]Alston [1985], pp. 83-84.

Well, no, or at any rate not yet. Coherence requirements might still hold if satisfying them is a way of satisfying some more fundamental norm that is directly in the business of promoting true beliefs? for instance, a goal norm like TB. More precisely, it might be the case that whatever doxastic state one ought to adopt in order to satisfy some truth-related goal norm, also has the property of satisfying some coherence requirement. In that case, coherence requirements would have no independent normative force; it would not be the case that one ought, for instance, to avoid holding contradictory beliefs simply *for the sake of avoiding contradiction*. Instead, satisfying coherence norms would be of instrumental importance.

One might, for instance, make the following simple argument in favor of NC:

1) TB holds: one ought to hold as many true beliefs and as few false beliefs as possible

2) Holding beliefs that satisfy NC (that is, non-contradictory beliefs) tends to result in holding more true beliefs and fewer false beliefs than does holding contradictory beliefs

3) Therefore, holding beliefs that satisfy NC results in one's also satisfying TB, while holding beliefs that violate NC does not

Conclusion One ought to hold beliefs that satisfy NC

Now, (2) is not obviously true, and (3) does not obviously follow from (1) and (2). Indeed, in Section III below I will criticize arguments of this sort. But I think that, if asked to explain *why* we think there's something wrong with holding contradictory beliefs, many of us might reply with something like this argument.[6] After all, if we hold a pair of contradictory beliefs, at least one of them is guaranteed to be false; so if true beliefs are what we're after, violating NC seems like a bad strategy.

---

[6] In the literature, Branden Fitelson, Kenny Easwaran and David McCarthy present an argument for NC that, while significantly more sophisticated (for instance, relying on a dominance avoidance norm rather than TB) roughly follows this pattern. See Fitelson, Easwaran and McCarthy [manuscript].

A more prominent use of the Instrumental Thesis is James Joyce's attempt to offer a purely epistemic, rather than pragmatic, argument for Probabilism (the claim that one ought not assign a credence function that violates the axioms of probability).[7] Joyce endorses a norm of truth for full beliefs along the lines of TB, and also endorses an analogous norm of gradational accuracy for credence functions, requiring that "an epistemically rational agent . . . must strive to hold a system of partial beliefs that, in her best judgment, is likely to have an overall level of gradational accuracy at least as high as that of any alternative system she might adopt."[8]

Joyce does not give a precise definition of gradational accuracy, which can be loosely characterized as the 'closeness' of one's credence values to the truth values of the propositions to which they are assigned (for instance, a high credence in a true proposition is more accurate than a low credence; a low credence in a false proposition is more accurate than a high one). Instead, he applies a number of axioms to set bounds on what he takes to be reasonable ways of measuring accuracy, and goes on to show that, for any of these reasonable accuracy measures and any credence function that violates Probabilism, there is another credence function that satisfies Probabilism that is more accurate (according to the specified measure) regardless of the truth values of propositions about which one is opinionated. Conclusion: one ought to satisfy Probabilism.

In my terms, Joyce endorses a goal norm of accuracy dominance-avoidance; proves that satisfying a coherence norm, Probabilism, is an effective means of satisfying the goal norm; and, by way of the Instrumental Thesis, concludes on that basis that one ought to satisfy Probabilism. The genius of his approach is that while satisfying Probabilism depends only on one's doxastic states, and accuracy depends both on one's doxastic states and the truth values of the propositions to which one assigns credences, Joyce seems to be able to show that having one's doxastic states arranged according to a certain pattern

---

[7]Joyce [1998], [2009]. See also Leitgeb & Pettigrew [2010 a,b] for an alternative attempt to derive Probabilism from accuracy norms.

[8]Joyce [1998].

always gets them closer to the truth, whatever the truth happens to be. Even if you don't know what the truth values are, and so cannot go about satisfying the goal norm directly, you can make progress by satisfying the coherence requirement.

Various objections to Joyce's argument are possible. Some are outside the scope of this paper; my purpose, after all, is not to determine whether Probabilism holds or not. Rather, I aim to figure out whether the relationship that Joyce claims to show between accuracy dominance-avoidance and Probabilism is the right *sort* of relationship to account for Probabilism's normative force, and relatedly, whether Joyce's argument is the right *sort* of argument to establish that Probabilism holds, if in fact it does. Certain objections to Joyce's argument are highly relevant to those questions, and will I turn to them in the next section.

I have not clearly stated just what sort of means-end relation must hold between, on the one hand, the doxastic states satisfying a coherence requirement, and on the other, a goal norm, in order for the Instrumental Thesis to apply to that coherence requirement-goal norm pair. I have not said, for instance, whether satisfying the coherence requirement must be a sufficient condition for satisfying the goal norm, a necessary condition, or both. When the Thesis is used to account for the force of an *unconditional* coherence requirement, that is, one that applies in any possible circumstances, then it must be the case that in any possible circumstances, if one does not satisfy the coherence requirement, then one also does not satisfy the goal norm.[9] So the Thesis will not offer a satisfactory explanation if adopting a coherent state is a merely sufficient condition for satisfying the goal norm, but there are incoherent states that would work just as well. On the other hand, it doesn't

---

[9]One might try to argue here that, on analogy with rule utilitarianism in ethics, that because satisfying a coherence requirement leads to satisfying a genuine goal norm *for the most part*, one should satisfy the coherence requirement even in cases of foreseeable conflict. I am admittedly unsympathetic to rule utilitarian approaches. However, I would suggest that the best version of such an argument, rather than claiming that (for instance) TB unconditionally establishes the force of NC, instead adopts an alternative goal norm along the following lines: "One ought to satisfy those requirements that, for the most part, result in. . . ." The structure of the instrumental relation is then as specified here. Section IV demonstrates the means of arriving at a case of conflict between even this 'rule utilitarian' goal norm and NC, though the considerations raised in Section VI suggest that such a conflict, occurring as it would only under very unusual circumstances, might not be such a great problem for our ordinary conviction in NC after all.

appear that in any given circumstances there must be some particular coherent state such that the adoption of *that particular* state is a necessary condition for satisfying the goal norm (the coherence requirement is vindicated if there are several ways of satisfying the goal norm, but *all* are coherent), nor yet that it must be the case that every single coherent state is sufficient to satisfy the goal norm.[10]

I would suggest, then, that the Instrumental Thesis succeeds for a given pair of a goal norm and a coherence requirement if, in any possible circumstances, there is at least one coherent state (that is, a state that satisfies the coherence requirement in question) the adoption of which is sufficient to ensure the satisfaction of the goal norm, and it is necessary that one adopt *some coherent state or other*. (Of course, this claim will only hold if circumstances are individuated in a sufficiently fine-grained manner as to settle which states satisfy the goal norm; so the possible circumstances at issue here must normally be understood as settling, at least, the truth values of propositions over which an agent is opinionated.)

Accepting this suggestion, however, is not strictly necessary for the purposes of my argument. The cases I will consider as challenges to the Instrumental Thesis involve circumstances in which it appears that *only* incoherent states are sufficient means to satisfying the goal norm. Thus even if one held that the coherence requirements could be validated be a version of the Instrumental Thesis requiring only that in all possible circumstances some coherent state be *among* the sufficient means for satisfying the goal norm, the challenges will still apply. Similarly, they would apply if satisfying a coherence norm need only increase the chance of one's satisfying the goal norm, in cases where whether or not doing so

---

[10]Kolodny [2007] argues that if there are *any* consistent belief sets that do promote true belief — that is, if there are any coherent states the adoption of which is not sufficient to satisfy the goal norm — then NC is not a genuine requirement. I disagree; I think a requirement can be genuine if one always ought to satisfy it, even if there are certain *ways* of satisfying it that are unacceptable. This is in part a terminological dispute, however, and I am largely sympathetic to the spirit of Kolodny's suggestion, as he goes on to try to work out what I would call an instrumental argument for NC and he calls an error theory. My goal here is in large part to describe the general conditions under which it is possible to show that, in order to satisfy a goal norm, one always ought to act in accordance with a coherence requirement, whether or not this is taken to establish the existence of the coherence requirement itself.

is a matter of objective chance.

3   A CHALLENGE TO THE INSTRUMENTAL THESIS: TWO CASES

This section will focus on a pair of cases that poses a *prima facie* challenge to the plausibility of applying the Instrumental Thesis to particular coherence requirements: the case of the mad scientist first mentioned in the introduction, and Michael Caie's objection to Joyce's argument for Probabilism. Each case works by describing circumstances in which violating a coherence requirement is the only way satisfying the goal norm from which it might have been thought to derive its force. These cases therefore present us with a trilemma: we must either adopt a different (and perhaps less plausible) goal norm; reject the coherence requirement in question, at least in its unconditional form; or deny that the Instrumental Thesis applies. In the next section, I will then identify the common mechanism at work in each case's challenge to the Thesis, and consider the extent to which those challenges can be generalized to apply to other pairs of goal norms and coherence requirements.

I provided one example of a conflict between a coherence requirement and a goal norm (namely, NC and TB) in the introduction: the example of the mad scientist who offers to improve your deductive powers, leading to many new true beliefs and/or correction of false beliefs, if only you can demonstrate to him that you hold contradictory simultaneous beliefs about whether seven is a prime.[11] Such a case presents a *prima facie* challenge to any attempt to use TB and the Instrumental Thesis to explain the force of NC, along the lines of the simple argument sketched in Section I.[12] That argument suggested that satisfying

---

[11]The case is inspired by an example originally due to Roderick Firth [1981], as related, and given illuminating discussion, in Berker [2013]. Firth's case does not explicitly deal with coherence requirements, though it does challenge the plausibility of norms like TB. Other similar cases can be found in Greaves [2013].

[12]This simple version of the case is appropriate to the simplicity of TB. As we will see, the case can be modified to apply to more nuanced goal norms. For instance, if one's preferred truth norm weighs true beliefs

NC is important because doing so results in holding more true beliefs and fewer false ones. Yet it looks like, in the mad scientist case, there is a possible incoherent state that includes more true beliefs and fewer false beliefs than any state that one can achieve by remaining coherent; namely the state that results from holding contradictory beliefs about the number seven, and thereby earning the mad scientist's reward. The challenge to the Instrumental Thesis claims that in this case, the ideal state picked out by TB is therefore an NC-violating one, and so in order to satisfy TB it is necessary to violate NC. But then it cannot be the case that NC stands in an appropriate instrumental relation to TB in every case, and so TB cannot explain the force of NC by way of the Instrumental Thesis.

Each horn of the trilemma suggests a possible response to this case. Perhaps NC really serves to promote a goal norm other than TB, and we should simply reject TB in favor of this alternative. (If cases like the mad scientist arose *only* for TB, this would be an obvious choice; the problem will be made more threatening once it becomes clear how many goal norms such cases threaten.) Perhaps in the mad scientist case holding contradictory beliefs is indeed the rational thing to do, and we should reject the unconditional NC (while still perhaps maintaining that avoiding contradiction is a good idea most of the time). Or perhaps both TB and NC hold, but NC does not derive its force through a means/end relation with TB.

The mad scientist case has various features that might lead to its being dismissed as a valid counterexample to the Instrumental Thesis as applied to TB and NC. In the mad scientist case, one's contradictory beliefs *causally* bring about the satisfaction of TB. Perhaps the means/end relationship involved in the best version of the Instrumental Thesis is a non-causal one. In the mad scientist case, one is enticed to hold contradictory beliefs about one topic (whether seven is a prime number) in order bring about improvements in one's deductive faculties, and thereby, true beliefs that will at least for the most part have

---

about some subjects more heavily than others, the mad scientist could offer the ability to deduce true beliefs about the very most important subjects. This point will be addressed in more detail in Section IV; for now, I only ask that readers not dismiss the case merely on the grounds that they are too sophisticated to have accepted TB in the first place.

nothing to do with numbers at all. Perhaps the most promising version of the Instrumental Thesis does not warrant this sort of trade-off, but does permit holding contradictory beliefs about a (perhaps quite specific) topic in order to promote truth in one's beliefs about that very topic. Finally, in the mad scientist case, any epistemic success that one achieves as a result of believing a contradiction comes about via an improvement in one's deductive capacities — but it might be argued that the demands of epistemic rationality include the demand that one carry out excellent or even perfect deduction.[13] So perhaps what epistemic rationality requires in this case is neither that you believe a contradiction, nor that you remain ignorant of the truths that improved deductive capacities might deliver; rather, it is that you satisfy NC while *also* arriving at the correct deductive conclusions.

As it turns out, none of these problematic features of the mad scientist case is necessary in order to generate the sort of conflict that can be used to challenge the Instrumental Thesis. Let us turn to an example that does not trade on a causal relationship between means and end, nor on a tradeoff between beliefs on different topics, nor on a dilemma that one could avoid by satisfying other plausible rational requirements. This case was proposed by Michael Caie [2013] as the centerpiece of his objection to Joyce's argument for Probabilism.

Recall the structure of Joyce's argument: one ought to satisfy Probabilism because, whatever the truth values of the world may be, doing so is an effective means to satisfying the goal norm of accuracy dominance-avoidance. What exactly does that goal norm amount to? Joyce spells out the norm as follows:

> **ADA**: One ought to assign a credence function c such that there does not exist any possible alternative function, c^*, over the same set of propositions, that is more accurate than c in every possible world.

---

[13]There are possible variants of the mad scientist case in which the mechanism through which the scientist offers his rewards is not an improvement in your mental capacities; he might instead, for instance, offer a great deal of new *evidence* from which you are able to deduce new truths. I will initially focus on the deductive-capacity version for expositional purposes, as it more clearly invites possible responses that I wish to consider; however, I will address a new-evidence version of the case in Section V.

Joyce provides an intricate proof that, if one assigns a function that satisfies Probabilism, then one also satisfies ADA; and that if one assigns a function that *violates* Probabilism, then one also violates ADA (that is, there does exist some alternative function that is more accurate in every world).[14] And on the face of it, this looks like a pretty good case for satisfying Probabilism. It's easy enough to feel the force of the intuition that it is a mistake to hold credence function c if some alternative, c^*, is guaranteed to be more accurate.

Caie argues, however, that we should not be so quick to embrace Probabilism; in fact, he claims, there are cases (albeit somewhat odd ones) in which it is epistemically rational to violate Probabilism. His central case is that of self-referential propositions like the following, designated P:

(P): John currently assigns to P a credence less than or equal to .5

Now, at any given time, P is either true or false. If I assign a credence less than or equal to .5 to P, then P is true (supposing I am the John referred to), while if I assign a credence greater than .5, or fail to assign any credence at all, then it is false. Suppose, for simplicity's sake, that my credence function assigns values to only two propositions, P and $\neg P$. Joyce's result tells us that if I assign credences that satisfy Probabilism (and that therefore my credences in P and $\neg P$ sum to 1) then there is no Probabilism-violating credence function that is more accurate in every possible world; if I assign credences that violate Probabilism, then there must be *some* Probabilism-satisfying function guaranteed to be more accurate (though which it is depends on the choice of scoring rule).

In particular, suppose I assign the following Probabilism-satisfying function C:

$C(P) = .5$
$C(\neg P) = .5$

Then (given the content of P) we can conclude that P must be true (since I assign a credence less than or equal to .5) while $\neg P$ must be false; my credence assignment,

---

[14]The details of Joyce's proof are not essential here, as neither Caie nor I calls into question the link between Probabilism and ADA.

therefore, is not particularly accurate. Now Joyce is perfectly correct that C satisfies ADA; there is no Probabilism-violating function that is more accurate than C regardless of the truth values of P and $\neg P$. Any such function that is more accurate (given a plausible scoring rule) than C when P is true, must be less accurate than C, according to that same scoring rule, when P is false. But Caie's crucial insight is that, so long as I assign function C, P *cannot be false.* P is false only when I assign either a credence greater than .5 to P, or assign no credence to it at all. But both of these states of affairs are inconsistent with my assigning C.[15]

Here then is a Probabilism-violating function that, *so long as I assign it*, is certain to be more accurate than C:

$$C^*(P) = .5$$
$$C^*(\neg P) = 1$$

$C^*(P)$ is, of course, equal to $C(P)$. On the other hand, $C^*(\neg P)$, unlike $C(P)$, is *perfectly* accurate whenever I assign $C^*(P)$; assigning $C^*(P)$ guarantees that $\neg P$ is true, and $C^*(P)$ assigns a credence of 1 to it. It follows that $C^*(P)$ is more accurate than C according to any scoring rule with the following property:

> If any two credence functions c and c^* assign the same values to all propositions other than p, and c^* assigns a higher value to p, then c^* is more accurate than c at any world in which p is true and less accurate than c at any world in which p is false.

But Joyce requires that all his candidate scoring rules have this property.[16] And so, at any world where $\neg P$ is false (and in particular, any world where I assign either C or C^*), C^* is more accurate.

Of course, Joyce is not committed to the claim that C must, in virtue of satisfying Probabilism, be more accurate than C^* in every possible world. Instead, his argument

---

[15]Caie takes himself here to be applying a familiar principle from practical decision theory: roughly, dominance reasoning is only properly applied so long as acts and states are *independent*, in the sense that an agent's choice of act does not affect the expected outcome of the act.

[16]See for comparison the Dominance axiom in Joyce [1998], p. 593.

establishes only there is some single Probabilism-satisfying function that is more accurate than C^* in every possible world, and that there is no Probabilism-violating function that is more accurate than C in every possible world. But consider what it would take for any credence function to be more accurate than C^* in a world where, as in all the cases we've been considering, P is true and $\neg P$ is false. Coming up with such a function is not difficult:

$$C^{**}(P) = 1$$
$$C^{**}(\neg P) = 0$$

C^{**} is, in fact, *perfectly* accurate at such a world: the credences it assigns perfectly match the truth values. But if I, the 'John' referred to in P, were to assign C^{**}, then I would no longer be in such a world. Given the content of P, it would then be the case that P would be false and $\neg P$ would be true — and so C^{**} would be perfectly *inaccurate*. My assigning any particular credence to P has the potential to change P's truth value. And so while C^{**} may be an excellent choice for other agents who are concerned with promoting their accuracy, in so far as I want to promote my accuracy it's a terrible choice for me.

As it turns out, my adopting C^* results in a better accuracy score for me than does adopting any Probabilism-satisfying function. And it does so foreseeably; the superior accuracy of $C*$ does not rest, for instance, on empirical facts that I might or might not be in a position to know. As long as I understand the content of P, I am in a position to be certain of what accuracy score I will receive, on any given measure, for any function over P and $\neg P$ that I might adopt.[17]

None of this in any way challenges the claim that I will satisfy ADA iff I adopt a Probabilism-satisfying function. But it *does* show that only by adopting a Probabilism-violating function can I satisfy the following alternative goal norm:

CADA: One ought to assign a credence function c such that there does not exist

---

[17]Of course, actually being certain of what accuracy score I will receive requires being opinionated about propositions other than P and $\neg P$. Understanding the content of P arguably does as well. C and C^* may unproblematically be extended to include the required credences, or supplemented with alternative doxastic states such as full beliefs, as long as the same changes are made in each case.

any alternative function, c^*, over the same set of propositions, the assignment of which is guaranteed to result in better accuracy than the assignment of c.

Although Caie himself does not put his challenge to Joyce in quite these terms, we may understand him as arguing as follows.[18] The plausibility of ADA as a goal norm is a result of conflating it with CADA. On learning that Probabilism-satisfying credence functions satisfy ADA, we imagine that so long as we satisfy Probabilism, while we don't rule out the possibility that some alternative credence function is *in fact* more accurate than ours, we do at least rule out the possibility that some alternative credence function that we might adopt is *guaranteed* to be more accurate. It seems plausible that adopting a credence function when there is a guaranteed more accurate alternative available constitutes a failure of rationality, and so ADA strikes us as a reasonable epistemic goal norm. But in fact, ruling out that sort of guaranteed-better-alternative requires satisfying CADA, not ADA. And far from satisfying Probabilism's being a foreseeably effective means of satisfying CADA, there are circumstances in which violating Probabilism is *necessary* in order to satisfy CADA.

Our predicament then is this: satisfying a particular coherence requirement, Probabilism, is always a means to satisfying one candidate goal norm, ADA. But it isn't clear why we should care about satisfying ADA, at least once we clarify how it differs from CADA; doing so doesn't seem obviously to be constitutive of being an ideally epistemically rational agent, nor does it seem like a goal we're committed to qua believers. CADA, in contrast, is an intuitively plausible goal norm, in that asserting it seems like a plausible way of spelling out the familiar epistemic evaluative point of view. But while in many circumstances satisfying CADA requires satisfying Probabilism, in some cases it foreseeably requires *violating* Probabilism instead. We return to our trilemma. We may insist that, despite its flaws, ADA and not CADA is the correct goal norm, and thereby preserve Probabilism; we may

---

[18]For a similar reading, see Carr [unpublished].

abandon ADA in favor of CADA while insisting on the Instrumental Thesis, losing Probabilism in the process; or we may hold on to both Probabilism and CADA, but deny that Probabilism derives its force from CADA via the Instrumental Thesis.

4 GENERALIZING THE CASES: THE POSSIBILITY PROBLEM

The previous section presented *prima facie* challenges to two potential applications of the Instrumental Thesis. The mad scientist case challenges the Thesis's ability to account for NC in terms of TB. Caie's case similarly challenges the Thesis's ability to account for Probabilism in terms of CADA; and, since Caie sees CADA as the most plausible accuracy dominance avoidance norm, the case challenges by extension accuracy dominance arguments for Probabilism generally. As things stand we face a trilemma, and we have yet to consider the merits of each horn. Before doing so, however, it is worth trying to figure out the essential features of each of these cases, those which drive their ability to challenge the Instrumental Thesis. Doing so serves two purposes. First, it will shed light on the generality of the challenge, and so on the range of additional applications of the Thesis that we might expect to run into similar problems. Second, it will help us figure out what an adequate response to the challenge must look like.

Let's begin by considering just the mad scientist case. In particular, for simplicity's sake, take a version of the mad scientist case in which whether or not one believes a contradiction, and thereby earns the scientist's reward, does *not* affect the truth values of any of the propositions over which one is opinionated.[19] The mad scientist case is supposed to be one in which TB and NC conflict, because accepting the scientist's offer (and thereby violating NC) results in holding more true and fewer false beliefs than any other possible belief set. TB tells one to adopt the belief set that maximizes true belief and minimizes false belief, from among the possible options. Notice: the mad scientist case works only if the

---

[19]Mad scientist cases may of course be described in such a way that the relevant truth values do depend on one's doxastic state. However, no such effect is essential to the challenge they pose to the Instrumental Thesis; in contrast, it is essential to the Caie-like cases discussed below, and so I will delay discussion of this effect and consider it in the context of those cases.

belief set resulting from violating NC is possible, and any alternative belief set containing more truths or fewer falsehoods is *not* possible.

Independently, each of these claims might seem plausible. Despite whatever psychological barriers there might be to believing a contradiction, there is a sense in which doing so is possible; at the least, it's metaphysically possible, and we can evaluate counterfactuals with antecedents like "If I believed that seven is prime and that it is not . . . ." On the other hand, there is a *sense* in which adopting a belief set containing more truths and fewer falsehoods than that offered to me by the scientist is impossible; as stipulated in the case, any such belief set is not one that I'm capable of reasoning to, given my actual evidence and actual mental abilities. The crucial point, however, is this: neither of these senses of 'possible,' these forms of modality, *both* counts the set resulting from the scientist's offer (which I can't reason my way to, given my actual limitations) as possible, *and* counts all sets containing more true and fewer false beliefs (for instance, the metaphysically possible belief set excluding the belief that seven is not a prime, but otherwise identical to the set resulting from the scientist's offer) as impossible. If we are presented with an argument that purports to justify NC in virtue of a version of TB that employs metaphysical possibility (for instance), then whatever the argument's independent merits, a mad scientist case cannot serve as a counterexample: while the belief set resulting from accepting the scientist's offer will count as possible, so will the belief set that includes all and only the truths, and so obviously the mad scientist's set will not be the *best* among the possible options. The mad scientist can't offer you anything better than the best metaphysically possible option. If instead a sense of 'possible' is used that applies only to belief sets that I am capable of reasoning my way to given my actual mental capacities, then while no option *better* than the scientist's offered belief set is possible, that every set isn't possible either.

We might then more appropriately describe TB not as a goal norm, but rather as a family of closely related goal norms. We can think of a fully specified version of TB as having two components that work together to pick out the belief set that one ought to

adopt. First, TB provides a preference ordering among (broadly logically possible) belief sets, according to the number of true and false beliefs that they include; sets with more true and fewer false beliefs rank higher. Second, a particular version of TB must specify the set of 'possible' belief sets, such that one ought to adopt whichever belief set from that set of sets ranks highest in the preference ordering. So when we wish to give TB definite content, we must specify what kind of modality is invoked by the use of 'possible' in the norm.[20] And many ways of so specifying TB will result in the verdict that, in the mad scientist example, TB and NC do *not* conflict, for they conflict only given a sense of 'possible' that rules in the scientist's offer but rules out the otherwise preferable alternatives. Of course, there are kinds of modality that will do the trick, and treat mad scientist cases as genuine instances of TB-NC conflict. If nothing else, we can specify a sense of 'possible' in an extensional manner, explicitly stipulating the possibility or impossibility of each belief set to ensure that we end up with a conflict; in so doing, we will be identifying a *version* of TB that cannot be used, along with the Instrumental Thesis, to justify NC. But insofar as we are interested in whether NC is in fact validated by a genuine goal norm, the mere fact that there exists some candidate, involving a contorted sense of possibility, that fails to validate it is not particularly worrisome. If someone wished to use the mad scientist case alone to show that *no* genuine goal norm in the TB family justifies NC, they would first owe us an argument that, if there is any genuine norm in the TB family, then it invokes one of the particular senses of possibility on which the mad scientist objection succeeds.

Mad scientist cases can be adapted to challenge goal norm-coherence requirement pairs other than TB and NC. For instance, we can describe a mad scientist case in which

---

[20]Of course, there are ways of stating goal norms, and even of restating TB, that do not use the *word* 'possible.' What all goal norms must do is (explicitly or implicitly) specify the range of options from among which the goal norm counts as ideal whichever is highest in the preference ordering. One common way of carrying out this specification is an appeal to possibility, as found in my original statement of TB and the narrow-scope restatement in Section I; in this case, the question of how the goal norm specifies the range of valid options amounts to the question of how the use of 'possible' in the norm is to be understood. For simplicity, I will sometimes speak of the 'sense of possibility' used in a goal norm as a proxy for the way in which we should understand it as constraining the range of valid options, even in cases where the most natural statement of the goal norm may not in fact use 'possible.'

one is offered a highly accurate credence function in exchange for a violation of Probabilism, or a belief set that is exceptionally well supported by one's evidence in exchange for a violation of Deductive Consistency. One could be offered true beliefs about topics with which one is particularly concerned, or the correction of false beliefs about objectively important subjects. Essentially, whatever feature of a doxastic state is taken to be of fundamental concern (epistemically speaking), we can describe a case in which the mad scientist offers the means to achieve a doxastic state that does extremely well in terms of that feature, if only one will violate some particular coherence requirement. There is nothing about TB and NC in particular that makes them uniquely suitable for generating mad scientist-type challenges.[21] Nor must the mad scientist offer his reward in the form of an improvement in your mental abilities; he might instead, for instance, offer you a great deal of new *evidence* in exchange for violating a coherence requirement. However, any mad scientist case will, like the original, require that whatever goal norm is being considered uses a sense of possibility on which accepting the scientist's offer results in a possible doxastic state, but there is no possible way of getting to a preferable state without accepting his offer. Mad scientist cases are thus highly *general* in the sense that they do not depend for their force on the specific content of the goal norm or coherence requirement to which they apply, but highly *specific* in that they only apply to goal norms involving one of a narrowly constrained range of forms of modality.[22]

In the discussion above, I considered only instances of mad scientist-like cases in which the agent's choice of doxastic states did not affect any of the truth values over which

---

[21]There is one limitation on the range of goal norm-coherence requirement pairs to which mad scientist cases may be applied. The doxastic state required by the goal norm in question must depend on the agent?s circumstances — as, for instance, which belief set satisfies TB depends on which propositions are true, but which credence function satisfies ADA does not.

[22]There is a special case in which the sense of possibility employed is irrelevant: that in which the goal norm in question simply explicitly requires that one adopt a state satisfying some coherence requirement. For instance, if the goal norm is a narrow-scope requirement that one adopt consistent beliefs, then no matter what belief sets are considered possible, it will never be the case that the possible belief selected by the goal norm is an inconsistent one. However, adopting such a goal norm is equivalent to abandoning the Instrumental Thesis altogether; it is to claim that being coherent matters because of the final value of coherence, rather than because of the instrumental relation between coherence and some other normative requirement.

she was opinionated. Of course, strictly speaking, the doxastic state that one adopts is always among the facts of the world; there are some propositions that have different truth values depending on whether or not one accepts the mad scientist's offer. Cases like Caie's exploit this fact. Let us now turn from considering cases like the mad scientist case, to cases like Caie's, and see how the insights yielded by the former carry over to the latter.

Caie's case is meant to show that, while there are no circumstances in which satisfying ADA requires violating Probabilism, there are circumstances in which satisfying CADA requires violating Probabilism: namely, cases in which one is opinionated over certain sorts of self-referential propositions. Joyce's argument for Probabilism works by showing that, for each logically possible Probabilism-violating credence function, there is a logically possible Probabilism-satisfying function that is more accurate at every logically possible world. Just as the number of true beliefs that one holds depends both on what one believes and on what is true, one's accuracy depends both on one's credence function and on the truth values of the propositions over which one is opinionated; that is, accuracy is a function of a pair consisting of a credence function and a truth value assignment (henceforth TVA). Joyce tells us that for every Probabilism-violating function c, there is a Probabilism-satisfying function c^* such that, for any TVA T, there is a *possible* pair of c^* and T that receives a higher accuracy score than does the pair of c and T, with *possible* construed in such a way that it applies to any pair of a logically possible credence function and logically possible TVA.

In objecting to Joyce's choice of ADA as the correct dominance-avoidance goal norm, Caie is really objecting to Joyce's implicit decision to adopt this particular sense of possibility, for the purpose of specifying the comparison class that determines whether a given function is non-dominated. In switching from ADA to CADA, Caie is essentially telling us that the dominance-avoidance norm we *should* care about is one that treats as possible only those pairs of a credence function and a TVA such that the TVA's obtaining is consistent with the agent in question actually adopting the credence function; that is, the possible

credence function-TVA pairs are those pairs that are instantiated at metaphysically possible worlds. Caie's constraint may not immediately seem suspect since it does not rule out any metaphysically possible worlds, but it is in the first instance a constraint on credence function-TVA pairs rather than on possible worlds, and it *does* importantly rule out a number of those pairs. Once we adopt a reading of the dominance-avoidance norm that uses this sense of 'possibility' to pick out the relevant class of credence function-TVA pairs (by requiring, not that there be no alternative function that is more accurate evaluated relative to every possible world, but that there be no function that is more accurate evaluated relative to every TVA consistent with its adoption by the agent under consideration), Joyce's claim that any Probabilism-violating function is necessarily accuracy-dominated by some Probabilism-satisfying function comes out false in just those cases Caie points to. We can interpret CADA as a version of ADA modified to make use of this more restrictive sense of possibility.[23]

As in the mad scientist cases, then, Caie's ability to generate cases of conflict between a goal norm and a coherence requirement depends on how he interprets the sense of possibility at work in the goal norm. On the sense of possibility Joyce relies on, which credence functions satisfy ADA for an agent is independent of that agent's circumstances, and every ADA-satisfying credence function is a Probabilism-satisfying one; it follows that there are no circumstances in which ADA and Probabilism come into conflict. On some senses of possibility more constrained then Caie's, where it takes more for a credence function-TVA pair to count as possible for an agent than for there to be some logically possible

---

[23]Rendering the requirements of CADA fully explicit, so as to show how it makes use of this sense of possibility, is a somewhat complicated task. Where ADA requires that we adopt a credence function such that no alternative is more accurate at every possible world, CADA requires that we adopt a credence function such that no alternative function is *guaranteed* to be more accurate. Let us say that a function c^* is guaranteed to be more accurate than c for agent A iff for every TVA T that obtains at at least one metaphysically possible world where A assigns c, 1) there is at least one metaphysically possible world where T obtains and A assigns c^*, and 2) the pair consisting of T and c^* is more accurate than the pair consisting of T and c. Whether a credence function counts as ideal under CADA, then, depends not on the relative accuracy of all credence function-TVA pairs that do and do not include that particular function, but rather on the relative accuracy of the pairs that include the function *and are instantiated at a metaphysically possible world*, and the pairs that don't include a different function instead *and are instantiated at a metaphysically possible world*.

world at which the agent adopts the function and the assignment obtains, the Probabilism-violating credence functions he advocates would not count as possible. It is only by choosing precisely the sense of possibility that he does, in which the set of possible credence function-TVA pairs depends on an agent's circumstances and some circumstances rule *in* his preferred Probabilism-violating option while ruling *out* alternatives that are still more accurate, that Caie is able to generate conflicts between Probabilism and accuracy dominance avoidance.

This is not to say that Caie has made a mistake. Perhaps the sense of possibility that he relies on is the sense at work in the most plausible accuracy dominance-avoidance norm. My point at present is only that the success of Caie's challenge to the Instrumental Thesis, as applied to accuracy dominance and Probabilism, is contingent on the particular sense of possibility that he builds into CADA. His argument should only convince us of a more general failure of accuracy dominance arguments to justify probabilism if we independently accept that the most plausible accuracy dominance-avoidance goal norm is one using this sense of possibility.

As with mad scientist cases, though, the force of the Caie case does *not* depend essentially on its application to an accuracy-based goal norm, or to Probabilism. Rather it depends on his adopting a sense of possibility such that which doxastic state-TVA pairs count as possible is in part a matter of the agent's circumstances, and then specifying circumstances in which some Probabilism-violating function counts as possible while any more accurate Probabilism-satisfying function (that is, any function higher in the preference ordering established by ADA) counts as impossible. Once we choose a goal norm on which the set of pairs of doxastic state and TVA that count as possible for an agent depends on that agent's circumstances, then we are in a position to try to construct a Caie-like case of conflict between that goal norm and a given coherence requirement: simply stipulate circumstances in which adopting some incoherent state brings about a state-TVA pair high in the goal norm's preference ordering, and any higher-ranked pairs of a coherent state and

a TVA count as impossible (for instance, because the agent's adopting the state is inconsistent with the TVA's holding).[24] Mad scientist cases are really just the special case of this phenomenon in which, since the TVA is held constant among all possible outcomes, we may speak of the goal norm as treating doxastic *states* rather than *state-TVA pairs* as possible or impossible. In mad scientist cases only one TVA is possible, so a state is possible iff the pair consisting of the one TVA and that state is possible.

---

[24]It is even possible to describe cases with much the same general structure as Caie's, but that rely on causal (rather than semantic) connections between the agent's doxastic state and the TVA — for instance, cases in which one's high confidence that one will succeed in some undertaking makes it the case that one will fail. See Greaves [2013] for discussion. Such cases have the advantage of greater real-world relevance than Caie's; nonetheless I have chosen to focus my discussion on cases involving self-referential propositions for (relative) clarity of exposition.

5 CONFLICT-IMMUNE GOAL NORMS

Suppose we accept that cases like those discussed in the previous two sections count decisively against applying the Instrumental Thesis to at least some particular goal norm-coherence requirement pairs: those in which the goal norm involves the wrong sort of possibility. What then are we to say about that status of coherence requirements? One option is to deny that there are any genuine unconditional coherence requirements. A second option is to maintain the existence of some such coherence requirements, but to deny that they derive their force from goal norms in the manner described by the Instrumental Thesis. These two possibilities will be considered in Section VI, below. The present section will discuss a final possibility: that one or more unconditional coherence requirements derive their force from an instrumental relation with goal norms *not* subject to cases of conflict like those described above.

If we aim to vindicate both some unconditional coherence requirements and the Instrumental Thesis, then we face the burden of coming up with some better goal norms, goal norms that can be used to justify the coherence requirements to which we are committed without giving rise to cases of conflict like those discussed in sections III and IV.

In and of itself, this is not actually such a difficult task. Indeed, we have already come across one goal norm for which no such conflicts can arise: Joyce's ADA. Remember, Caie's objection to Joyce's argument for Probabilism was not that ADA conflicts with Probabilism, but rather that ADA is not a goal norm that we ought to be concerned with in the first place. To make the general point more explicit: it is straightforwardly true that, evaluated relative to any world, the logically possible belief set that will contain the most truths and fewest falsehoods will be one that assigns belief to only propositions true at that

world; assuming classical semantics, this belief set will satisfy NC. Similarly, it is the case that, evaluated relative to any world, the most accurate of the logically possible credence functions will be the Probabilism-satisfying credence function that assigns values of 1 to all true propositions and all false propositions. In order to arrive at a truth-based goal norm that will never come into conflict with NC, therefore, we must simply choose a norm that requires us to adopt the belief set containing the most truths at our world, *out of all logically possible belief sets*. We can arrive at an accuracy norm that will never conflict with Probabilism by choosing one that requires us to adopt the credence function with the highest accuracy, *out of all logically possible credence functions*.

These goal norms, however, aren't what we're looking for. For one thing, they essentially require that rational agents be omniscient. There may be a perfectly good sort of epistemic evaluation that we sometimes engage in, in which we are concerned only with how close some agent comes to omniscience, and fault her for any false belief or less than perfectly accurate credence. But one can be rational without being omniscient, and so the observation that one must be coherent in order to be omniscient hardly amounts to an argument that one must be coherent in order to be rational. When we ask whether an agent is believing *rationally*, we are asking whether she meets some standards short of being such that there is no superior doxastic state that is metaphysically possible for her. Moreover, Caie's argument shows us that sometimes omniscience isn't even metaphysically possible; certain truths just cannot be believed by certain agents (while retaining their status as truths). So even were we otherwise disposed to think that requiring a rational agent to believe all the truths is not too high a standard for rationality, we might hesitate to demand that she do the metaphysically impossible.

The sort of evaluation that holds agents to standards tantamount to omniscience, moreover, is not the kind underlying ordinary appeals to the Instrumental Thesis. Joyce, after all, does not frame his argument for Probabilism by noting that if you had *perfectly* accurate beliefs, you would also have probabilistically coherent beliefs, and therefore you

are defective if you violate Probabilism. If that were his point, there would be no need to engage in a dominance-based argument in the first place. Instead, he explicitly argues that rational agents will 'hedge their bets' by assigning non-extremal credences (which is to say, credences that are not perfectly accurate in *any* possible world), and motivates the case for Probabilism with the claim that a rational agent must adopt the state that 'in her best judgment' will result in high accuracy.[25] I think we should read Joyce as arguing that an agent who violates Probabilism has failed to do the best, epistemically speaking, that she is equipped to do. As we have seen, the argument runs into trouble because it requires adopting a sense of possibility even *more* demanding than that underlying a demand for omniscience; but in Joyce's preference for an accuracy-dominance argument over an accuracy-maximization argument, we can see his attempt to ground a standard of rationality in what is possible for an agent who is still capable of false beliefs and inaccurate credence assignments.

The real challenge of specifying suitable goal norms, then, is not simply to come up with norms that, if they hold, can support coherence requirements via the Instrumental Thesis. We can do that anytime. Nor even is it to come up with goal norms that can do the job and that are genuine parts of evaluative standpoints that we may, from time to time, adopt; sometimes we really are simply in the business of judging how far an agent has succeeded in believing truths. Instead, the hard problem is to come up with goal norms that can support coherence requirements *without* abandoning their claim to capturing the ordinary epistemic point of view from which we try to figure out the extent to which an agent's doxastic state counts as rational.

A promising strategy, I think, is to consider goal norms that constrain the range of 'possible' doxastic state-TVA pairs by excluding pairs that are instantiated (that is, the state is held by the agent under evaluation, and the TVA obtains) only at worlds that are in one way or another objectionable . . . for instance, worlds in which the agent's doxastic

---

[25]Joyce [1998], p. 579.

state was arrived at through certain prohibited processes. The intuitive idea here is that while the *goal* of rational belief may be (for instance) to maximize the number of true beliefs held, not just any way of acquiring true beliefs is a way of satisfying the goal norm. Compare the game of checkers. The goal of a player in a checkers game is to remove all of his opponent's pieces from the board. However, this does not mean that one can succeed at checkers just by reaching out and sweeping the opposing pieces onto the floor at any time; rather, one must achieve the goal by way of the sorts of moves treated as valid by the rules of the game.[26] If there are restrictions on the causal histories of worlds of instantiation that result in sets of state-TVA pairs for which conflicts between goal norms and coherence requirements cannot be generated, then we can adopt goal norms that maintain truth (or accuracy, or evidential fit, or whatever one's preferred fundamental epistemic value may be) as the goal of epistemic rationality, while also building in ground rules such that violations of coherence requirements will never count as 'winning moves.' So we can acknowledge that accepting the mad scientist's offer is a way of gaining true beliefs, without being committed to the claim that it is also a way of satisfying (the best version of) TB.

Note that this approach (in contrast to an alternative considered in Section VI below) need not rely on the claim that, for instance, satisfying NC is constitutive of belief, or satisfying Probabilism is constitutive of assigning credences. That is, we need not answer the epistemic analog of the question: is the 'checkers' player who refuses to abide by the rules, and knocks his opponent's pieces to the floor, *really* playing checkers at all? We can simply say: he is, at least, not playing checkers *well*. The rules are not built into the definition of the activity, but instead into the evaluative standards that we use to judge

---

[26]One might object here that the goal of checkers is not, strictly speaking, to remove the opponent's pieces from the board, but rather to *capture* those pieces, and that no one would think that knocking them onto the floor constitutes capturing them. But I think this is in fact precisely analogous to the epistemic situation. We can describe the goals of checkers from an 'external' perspective ('remove the pieces') or from an 'internal' perspective, couching our description in the language of the game's rules ('*capture* the pieces'). The 'capturing' description builds in the constraints on appropriate means. In soccer, we aim both to get the ball into the goal, and to *score* goals. Similarly, we can describe a truth norm from an external perspective ('maximize true beliefs') or an internal one ('maximize true beliefs, in the acceptable ways').

success. We are free to admit that it may be possible to believe while violating NC, while still denying that one can believe *well*, as evaluated from the ordinary epistemic point of view, while violating NC.

What then might be the particular constraints that make up the ground rules for the epistemic evaluative point of view? One possible response is that one should never assign doxastic attitudes to propositions like Caie's in the first place — tricky, self-referential propositions to which it is metaphysically impossible (for a particular agent) ever to assign true beliefs or perfectly accurate credences. Instead, the appropriate response in such cases is to suspend judgment, believe neither the proposition nor its negation, refuse to assign any credence, etc.

If you think this is the right response, you have it easy. Simply pick out any feature of such propositions, and exclude from the range of possibilities considered by one's goal norm any state-TVA pair only instantiated at worlds in which one assigns a doxastic attitude to such an objectionable proposition. If one then interprets the 'possible' in one's goal norm as applying to any state-TVA pair instantiated at any metaphysically possible world, *except* the worlds just explicitly ruled out, then it will not be possible to generate Caie-like cases of conflict between the goal norm and a given coherence requirement. But it will also not be possible to generate mad scientist-type cases. Recall that such cases only work when the sense of possibility used is one much more limited than metaphysical possibility, so that the mad scientist can offer you a result that's better than the best result possible through any other means.[27] The mad scientist, of course, can't offer you a result better than the best metaphysically possible one. So in modifying Caie's preferred sort of possibility to rule out Caie-like counterexamples, one blocks mad scientist counterexamples for free —

---

[27]I am assuming here that the goal norm and coherence requirement are such that whatever *metaphysically* possible state best satisfies the goal norm, tricky Caie propositions excluded, is coherent. Strictly speaking, that need not be the case: if we tried to use TB to justify Anti-NC, a coherence requirement satisfied only if one has at least one pair of contradictory beliefs, then even if the mad scientist offered to bring about the metaphysically best belief set-TVA pair, the one in which all and only the truths are believed, accepting the offer would result in violating Anti-NC. But surely no one would have attempted to use TB and the Instrumental Thesis to justify Anti-NC in the first place, and indeed I can think of no plausible application of the Instrumental Thesis for which the assumption I make here fails.

though one still faces the objection that the sense of possibility in play here, without further modification, leads to overly demanding goal norms, even if they are able to avoid conflict with coherence requirements. In most cases, goal norms using this sort of possibility will make precisely the same demands as goal norms that require rational agents to adopt the best metaphysically possible doxastic state, so much of the reason for resisting *those* goal norms will carry over.

If, on the other hand, one does not think that all worlds in which one is opinionated over Caie-like self referential propositions can simply be ruled out *ex ante*, then matters are somewhat more complicated. We can still identify other senses of possibility that block both the Caie- and mad scientist-type cases so far considered. Recall that the original mad scientist case depends for its force on treating the state that results from accepting the scientist's offer as possible, but treating the state containing all the same beliefs that one could deduce with the help of the scientist's drug, but *not* containing the false belief about the number seven required in order to convince him to give you the drug in the first place, as impossible. If we adopt a version of TB that violates either one of these criteria — that either treats the scientist's offered state as impossible, or the slightly-better state as possible — then the case will not be an instance of conflict between NC and TB.

Of course, there's more than one way of ending up with a given belief set. One way of ending up with the slightly-better set of beliefs is by simply *guessing*, and getting very lucky. After all, even without the mad scientist's aid, there's nothing *stopping* you from holding those particular beliefs that just happen to be the ones he would help you arrive at. It seems plausible enough that a belief set should not count as the one that you ought rationally to adopt just in virtue of the fact that if you happened to adopt it by guessing, you would end up with lots of true beliefs — if it did, then at least in cases not involving Caie-like self-referential propositions, one would be rationally required by TB not to hold *any* false beliefs, on the grounds that it's possible to arrive at a belief set containing only truths by guessing. On the other hand, one might also arrive at the same belief set (the

one slightly better than what the mad scientist offers) by deducing its contents *without* the aid of the scientist's drug. Of course, one's deductive abilities might in fact be sufficiently limited that one is not able to deduce one's way to that belief set, but it does not seem so perverse for a standard of rationality to say that one *ought* to do so, as for it to say that one ought to simply guess one's way to the truth.

Call a world in which you accept the mad scientist's offer a 'cheating world,' a world in which you arrive at a belief set by guessing a 'guessing world,' and one where you arrive at a belief set by working out what follows from your evidence a 'deducing world.' The mad scientist case treats as possible the state-TVA pair instantiated in the one cheating world, and the pairs instantiated in a subset of the deducing worlds — those worlds involving only the deduction that one is actually capable of carrying out. It rules out all state-TVA pairs only instantiated in guessing worlds, which as noted above, seems like a good thing. One way of defeating the mad scientist case is to simply adopt a sense of possibility that rules out the cheating world, and thereby the belief set offered by the scientist, while otherwise allowing the same set of possibilities assumed by the case. But this seems like an insufficiently demanding standard of rationality, since it says that you ought only believe as many truths as you yourself are capable of reasoning your way to, no matter how bad a reasoner you may be. A *different* way of blocking the mad scientist case is to adopt a sense of possibility that may or may not include the state-TVA pair instantiated in the cheating world, but that does include those pairs instantiated at *any* deducing world (or at least at a large enough subset of those worlds to include at least one pair that TB ranks ahead of that instantiated in the cheating world).

Intuitively, this option responds to the mad scientist case by saying: it may well be better to accept the scientist's offer than to remain as ignorant as you currently are. But neither is what you *ought* to do. What rationality demands of you is that you work out for yourself what follows from your evidence, just as you would do if you accepted the mad scientist's offer, but without holding contradictory beliefs about whether seven is a prime

39

in the process. It is unfortunate if you are not able to reason so well as rationality demands, but it is no strike against a norm's counting as a standard of rationality that it requires good reasoning.

A version of TB that can require a particular belief set in virtue of the fact that the beliefs contained in it can in principle be arrived at through perfect deduction will never give rise to the particular sort of mad scientist case in which the scientist offers an improvement in one's deductive capabilities, because a change in an agent's actual deductive capabilities will never result, given this goal norm, in a different state's counting as ideal. So long as 'deducing worlds' include only those that can be arrived at by strictly deductive reasoning, moreover, this version of TB is immune to Caie-type cases. While it might be possible to deduce in a Caie-type case that adopting inconsistent beliefs will maximize the number of truths and minimize the number of falsehoods believed, it will not be possible, or any particular proposition p, to deduce both p and $\neg p$. So although in a Caie-type case an agent might be aware that there is some metaphysically possible world in which she would hold more true beliefs in virtue of violating NC, it does not follow that violating NC is required by the version of TB under consideration.

Of course, deductive reasoning is not the only form of good reasoning, and it is plausible that the best version of TB, or any other goal norm, may sometimes require doxastic states that can be arrived at through good reasoning but not deductive reasoning alone. Depending on how inclusive an understanding of good reasoning one favors, it may be the case that a goal norm using a sense of possibility that ranges over all state-TVA pairs that one can arrive at through good reasoning will sometimes fall prey to Caie cases. To avoid this result, it must (at least) be the case that the inference from 'I would have more true beliefs if I believed p' to 'p' does not count as good reasoning. This seems like a reasonable constraint.[28]

---

[28]See Berker [2013] for discussion of other disadvantages of goal norms that do not include some such constraint, beyond their difficulties in accounting for coherence requirements (though Berker's target is goal- or value-based approaches in epistemology, of which the Instrumental Thesis is one, rather than those goal

40

The strategy I have been discussing, which responds to the original mad scientist case by requiring that agents adopt the best doxastic state that can be arrived at through good reasoning, is still potentially threatened by other variants of the mad scientist case. For instance, we might construct a case in which, rather than administering a drug that improves deductive capacities, the scientist simply rewards coherence-violating agents with a great deal of new evidence, from which they are able to reason using their ordinary deductive capacities. In some cases, it might be tempting to say that a rational agent ought to obtain for herself the evidence offered by the mad scientist, but this surely cannot be an appropriate demand in every case — that is, unlike in the original case, we cannot respond by saying simply that a rational agent ought to refuse the scientist's offer, but nevertheless end up working out all the true beliefs that he offers on her own. The only response consistent with our earlier strategy, therefore, is to exclude accepting the scientist's offer in a new-evidence case from the range of possibilities that may be required by TB. This will be an unacceptable result for anyone who judges that a rational agent ought to believe a contradiction in order to gain access to the scientist's new evidence, but that is a judgment I do not share. I am therefore happy to maintain that the sense of 'possible' at work in genuine goal norms will range over state-TVA pairs only instantiated in worlds resulting from an agent's carrying out excellent reasoning, but not pairs resulting from the agent's acquiring new evidence by any means necessary.

I have not, admittedly, provided a fully specified account or defense of a particular sense of possibility that is most appropriately read into epistemic goal norms. What I have done is to show what constraints the chosen sense of possibility must impose in order to block certain types of counterexamples; show that some senses of possibility that excel at blocking the entire range of counterexamples are nonetheless implausible because the standards to which we hold rational agents, should we adopt goal norms that make use of them, are either to strict or too lax; and argue that senses of possibility that count state-TVA

norms in particular); and for some plausible suggestions about the boundary between what I term good reasoning and cheating.

pairs as possible or impossible based on the features of the worlds in which they are instantiated block at least a wide range of counterexamples while also maintaining intuitively plausible goal norms. I think that such an approach offers an appealing strategy for saving the Instrumental Thesis without either abandoning unconditional coherence requirements or adopting wholly implausible goal norms. The details of an argument from a particular goal norm to a particular coherence requirement must still be filled in, but when such an argument is offered, an appropriate construal of the sense of possibility used in the goal norm may serve to save the argument from apparent counterexamples.

6 ALTERNATIVE RESPONSES TO THE CHALLENGE

If we do *not* accept that there are instances of genuine goal norms instrumentally justifying unconditional coherence requirements — if, that is, we are persuaded by the challenges described in Section III, but not by the response in Section V — then we face a choice between two alternatives. We may deny that there *are* any genuine, unconditional coherence requirements; or we may maintain the existence of such requirements, but deny that the Instrumental Thesis offers the correct account of their force. In this final section, I consider the merits of these responses to the problem.

Should we deny that there are any unconditional coherence requirements? The challenge described above exerts some pressure in this direction; we have seen that a number of seemingly promising attempts to justify particular such requirements using the Instrumental Thesis fail. For readers who were not particularly committed to the existence of genuine coherence requirements to begin with, this challenge may be sufficient, if not to conclusively rule out the possibility of such coherence requirements, then at least to place the burden of proof firmly on the shoulders of those who think they have a solid positive argument with which to establish some particular coherence requirement.

Others will maintain a commitment to coherence requirements, and continue to hope for arguments capable of validating them. I count myself in this latter camp, the members of which may still disagree over which particular requirements are supposed to obtain. For myself, it is difficult to shake the sense that there are *some* grounds for faulting any agent who violates either NC or Probabilism (which is not to say that such violations might not be all-things-considered justified).

A skeptic about unconditional coherence requirements will find it easy enough to explain away this intuition of mine. After all, they can readily enough agree that there are many instances in which violating NC or Probabilism *is* criticizable. The cases of conflict I have discussed above are hardly ordinary; most of us neither encounter contradiction-loving mad scientists nor consider tricky self-referential propositions on a regular basis. So the skeptic might argue that apparent coherence requirements are in fact useful heuristics, which get us the right answer (say, by helping us satisfy a goal norm) when applied in most ordinary circumstances, but that potentially lead us astray in the sort of bizarre case that troublesome philosophers can dream up. Or there might be genuine norms of coherence with independent force, which when properly interpreted simply do not require that one be coherent under all possible circumstances. Or yet it might be the case that coherent doxastic states are a typical, though not constant, byproduct of satisfying distinct norms; for instance, it might be the case that if we reason successfully based on our evidence, the end result will typically be a coherent state.[29] Any of these states of affairs might trick our intuitions into perceiving *any* coherence violation as objectionable, as we overgeneralize from ordinary cases.

I do not deny that this is a live possibility. I can only report that, presented with the possibility, I am not immediately won over, but remain invested in trying to see whether there is any plausible way in which unconditional coherence requirements might be saved. Coherence requirements *seem* real, and so it is worth finding out whether they might *be* real. If no plausible account of them can be found, then perhaps falling back on this error theory will be the only reasonable response, but if some plausible account can be given then while it still might be that my intuition is in error, there will hardly be conclusive reason to reject it.

An alternative possibility, then, is that there *are* genuine, unconditional coherence requirements, but that they do not in fact derive their force from goal norms in the manner

---

[29]See Kolodny [2007].

44

described by the Instrumental Thesis. This possibility is consistent either with there *being* no genuine goal norms, or with goal norms and coherence requirements coexisting without the satisfaction of either standing in an instrumental relation to that of the other.

Suppose we conclude that there are no genuine goal norms, but that there *are* coherence requirements.[30] In carrying out epistemic evaluation of doxastic states, both our own and those of others, we would then consider whether those states were coherent in certain ways, that is, whether they instantiated certain patterns. We might also evaluate whether those attitudes satisfied some assortment of narrow-scope requirements other than goal norms; for instance, certain beliefs might simply be mandated for any agent in a particular evidential situation. But *none* of these requirements would hold in virtue of their relationship to a more fundamental norm expressing an epistemic goal, such as the promotion of true belief.

It would be difficult to recognize the evaluative standpoint just described as an *epistemic* one, or at least as our ordinary epistemic perspective. When we ask whether some doxastic attitude is epistemically required, or prohibited, we are typically not just asking whether there is *some* evaluative standpoint that yields such a verdict. After all, beliefs may well be required or prohibited from a prudential or moral point of view, but realizing that they are does nothing to answer the question of whether or not it is epistemically appropriate to hold them. Similarly, the fact that there is a possible evaluative standpoint that privileges coherence while not being at all concerned with the promotion of true belief does not mean that any such standpoint is among those we are concerned with when we ask whether coherence is epistemically required. Recall Alston's claim, quoted in Section II, that the epistemic point of view is *defined* by the aim of maximizing truth and minimizing

---

[30]Is this possible? Perhaps any coherence requirement entails a norm that technically satisfies my definition of a goal norm. For instance, if NC is true, then plausibly the following is also true: "If one is a rational agent, then one ought to satisfy NC." If you think that such a goal norm does follow from NC, then simply take the view under consideration to be one on which there are no genuine norms reflecting epistemic goals other than coherence.

falsity in a body of beliefs. We may be more liberal than Alston, and allow that a some-what wider range of goals might define points of view worthy of being called 'epistemic,' without thereby being required to place a viewpoint that involves no goal beyond formal coherence itself as on a level footing with those others.

In more concrete terms, what would be unsatisfactory about such an evaluative stand-point? Well, one advantage of tracing the force to coherence requirements to that of goal norms is that we can hope to provide a unified account of the force of multiple coherence requirements, rather than stipulating the importance of each in an ad hoc manner. NC and Probabilism, although applied to different doxastic attitudes, bear a strong resemblance to each other; each entails, for instance, that one should not be highly confident in the truth of inconsistent propositions. If the goal norm that justifies NC specifies the right way of promoting the truth for full beliefs, and the goal that justifies Probabilism analogously specifies the right way of promoting the truth for credences, then it is unsurprising that similar coherence requirements hold in each domain. If on the other hand one wanted to hold on to both NC and Probabilism while denying the existence of any goal norms, then their similarity would, implausibly, seem to be a matter of pure coincidence.

One strategy for attempting to *explain* the force of coherence requirements in the ab-sence of goal norms is to claim that these requirements are somehow constitutive of the attitudes to which they apply: for instance, NC is somehow constitutive of what it is for a mental state to be a belief, as opposed to say a supposition. This strategy, however, faces significant challenges. It would be relatively straightforward to claim that it is constitu-tive of beliefs that they must satisfy NC; that is, to deny that any two mental states, held by the same person at the same time, with contradictory contents are really beliefs. The claim might perhaps be rendered more plausible by amending it to state that one cannot *knowingly* or *intentionally* hold contradictory beliefs. But while all of these claims could account for the difficulty or impossibility of some ways of violating NC, they would not

say what is *wrong* with violating NC, should one manage to do so. A different sort of constitutive claim could indeed explain what is wrong with violating NC: the claim that it is constitutive of a mental state's being a belief that the state *ought* to satisfy NC, whether or not it actually does. But this approach runs into its own difficulties. Suppose that we are considering two mental states, each of which satisfy all the non-normative requirements to count as a belief. Indeed, the two states are identical except that one ought to satisfy NC, and therefore counts as a belief, while the other does not. Then learning that beliefs ought to satisfy NC is of no use whatsoever in figuring out which of these states ought to satisfy NC — we can only apply NC once we know which state is a belief, but we can only figure out which state is a belief by first figuring out which ought to satisfy NC. (If on the other hand there is always a *non*-normative as well as a normative one between beliefs and non-beliefs, then why not think it is the non-normative difference that determines which are beliefs?) This does not rule out the possibility that beliefs are constituted by the norms they ought to satisfy, but I think it does show that if beliefs are so constituted then it is a mistake to use terms like 'belief' to pick out the things to which norms apply.

Even should we overcome these difficulties sufficiently to pursue one of the constitutive strategies for justifying coherence norms, we might still ask whether such a strategy can plausibly deliver NC or Probabilism in particular. If it is to do so, I think the strategy we adopt must allow that it is at least sometimes possible for an agent to be in the state despite knowingly flouting the coherence requirement. For instance, it seems perfectly coherent to imagine an agent knowingly assigning a Probabilism-violating credence function, under the (perhaps false) belief that doing so was the best way of promoting accuracy. While I have some sympathy, for instance, for the claim that an agent who is utterly unconcerned with her *accuracy* doesn't count as assigning credences, the same simply doesn't seem true of an agent who is utterly unconcerned with her probabilistic coherence (except insofar as the violation of Probabilism foreseeably results in worse accuracy, but here we are considering the appropriate verdict on the assumption that no accuracy goal norm holds). Thus

I am more open to a strategy which takes aiming at accuracy or truth to be constitutive of doxastic states, than one which directly takes coherence to be so constitutive; but then we return to the difficulties, familiar from the earlier sections of this paper, for justifying coherence requirements in terms of their ability to promote truth or accuracy.

What if, rather than denying the existence of genuine goal norms, we suppose that there are some such goal norms, along with some coherence requirements, but that neither derives its force from the other? Denying that one derives its force from the other, of course, does nothing to prevent conflicts between them. So adopting an evaluative perspective according to which we are independently concerned with, say, satisfying both of Probabilism and CADA means there that the cases of conflict presented earlier will, rather than undermining the force of the coherence requirement, simply result in a dilemma in which we are genuinely caught between two constraints that cannot be simultaneously satisfied. Perhaps this is our epistemic position: being coherent and avoiding accuracy dominance are both important, but you can't always do both at once.

However, this mere addition of independent goal norms does nothing at all to *account* for the force of the coherence requirements, or to explain the similarities between the coherence requirements that seem plausibly to be applied to different sorts of doxastic attitudes. Moreover, it is hardly a silver bullet for our concern about locating coherence requirements within the ordinary epistemic perspective. While the result will be an evaluative standpoint that is concerned, in part, with the promotion of truth, it will also be the case that some doxastic attitudes will be required for reasons entirely independent of their truth, or prohibited for reasons independent of their falsehood. That is, while this sort of evaluative point of view is one which will endorse *some* doxastic attitudes for truth-related reasons, it may also endorse others for reasons that are *entirely* a matter of coherence. Perhaps this degree of concern with the truth is sufficient that we might accept the strategy in question as consistent with our ordinary epistemic point of view, but perhaps not. Finally, this strategy

raises the question of whether there is any principled way of weighing coherence require-
ments against goal norms, or whether there are simply cases in which the two come into
irresolvable conflict. A version of this strategy on which coherence requirements take lex-
ical priority over goal norms might deliver much the same result as my preferred approach
as outlined in Section V, on which coherence requirements derive their force from their
status as a means to satisfying a (suitably restricted) truth or accuracy goal norm; however,
it would lack the resources to *justify* the coherence requirements in instrumental terms. On
the whole then, I think that my approach remains a more appealing option, so long as we
can identify constraints that might be built into the goal norms that are sufficient to avoid
cases of conflict with coherence requirements, without themselves being too alien to our
ordinary epistemic point of view.

REFERENCES

[1] Alston, W. P. (1985). Concepts of epistemic justification. *The Monist*, pages 57–89.

[2] Berker, S. (2013). The rejection of epistemic consequentialism. *Philosophical Issues*, 23(1):363–387.

[3] BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge Univ Press.

[4] Caie, M. (2013). Rational probabilistic incoherence. *Philosophical Review*, 122(4):527–575.

[5] Carr, J. Epistemic utility theory and the aim of belief. Unpublished MS.

[6] David, M. (2001). Truth as the epistemic goal. *Knowledge, truth, and duty: essays on epistemic justification, responsibility, and virtue*, pages 151–69.

[7] Firth, R. (1981). Epistemic merit, intrinsic and instrumental. In *Proceedings and Addresses of the American Philosophical Association*, volume 55, pages 5–23.

[8] Fitelson, B., Easwaran, K., and McCarthy, D. Coherence. Unpublished MS.

[9] Greaves, H. (2013). Epistemic decision theory. *Mind*, pages 915–952.

[10] Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, pages 575–603.

[11] Joyce, J. M. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In *Degrees of belief*, pages 263–297. Springer.

[12] Kolodny, N. (2007). How does coherence matter? In *Proceedings of the Aristotelian Society*, volume 107, pages 229–263. Wiley Online Library.

[13] Leitgeb, H. and Pettigrew, R. (2010a). An objective justification of bayesianism i: Measuring inaccuracy. *Philosophy of Science*, 77(2):201–235.

[14] Leitgeb, H. and Pettigrew, R. (2010b). An objective justification of bayesianism ii: The consequences of minimizing inaccuracy. *Philosophy of Science*, 77(2):236–272.