RESPECT FOR TRUTH AND THE NORMATIVITY OF EPISTEMIC RATIONALITY

by

KURT L. SYLVAN

(Dissertation committee: Ernest Sosa (chair), Ruth Chang, Alvin Goldman, and Derek Parfit.)

CONTENTS

INTRODUCTION

Asking the right question is often the hardest part.

—Ted Sider[1]

1.   *Reasons, rationality, and problems of deontic significance: lessons from the practical sphere*

To ask the most important questions, epistemologists will need to acknowledge more distinctions than they have often done. On this score, I think they can benefit from reading ethicists and philosophers of practical reason, whose understanding of the high-level contours of normativity has tended to be more nuanced, at least in recent years. It has become increasingly common for these philosophers to distinguish

(1)   correctly responding to all the **reasons** that bear on whether to φ

from

(2)   φ-ing or refraining from φ-ing **rationally**.[2]

While drawing a distinction between (1) and (2) may sound picky in isolation, simple cases make it attractive. Consider a well-worn example from Williams (1981). It looks to Bernie as if his glass contains gin and tonic. Nothing in his circumstances indicates otherwise. As it turns out, some knave put petrol in his glass. That fact is a conclusive reason for Bernie not to take a sip. Still, it would be wrong to say that Bernie is *irrational* for planning to take a sip. He displays no lack of rationality. Rationality in its ordinary sense is a *narrow* notion, to use Scanlon's nice word: it is the positive counterpart of a strong, agent-oriented criticism we express with "irrational". Epistemologists would, I believe, be wise to consider a similar distinction. The distinction comes, I'll be arguing in Chapters 2 and 3, to a distinction between what beliefs would be epistemically **justified** and what we would be epistemically **rational** in believing. Indeed, this is an exact analogue of the distinction between (1) and (2), given a high-level view about justification I'll recommend in Chapter 1 that links it with reasons.

---

[1]Sider (2009: 384).

[2]To get a sense of how common this is, note that the distinction is explicitly drawn or taken seriously by Broome (1999, 2008, Ms), Dancy (2000), Kolodny (2005, 2008a–c), Lord (Ms₁), Parfit (2001, 2011), Raz (2004), Reisner (2011a), Scanlon (1998, 2007), and Schroeder (2009), among others. It is also presupposed by Kantian internalists about practical reasons—e.g., Korsgaard (1986), Markovits (2011) and Smith (1994). They claim that you have a reason to A only if you'd be motivated to A after undergoing "rational deliberation". If rationality *just were* responsiveness to reasons, this would be a patently circular view.

It is also worth noting that Donald Davidson favored a narrow understanding of rationality in his (1982) and (1984). While he did not write much about *normative* reasons, I think he would have been happy to contrast rationality with responsiveness to genuine normative reasons—though rationality would surely be tied to *motivating* reasons for him. The distinction is implicit in some passages in his early work: "The justifying role of a reason [...] depends upon the explanatory role, but the converse does not hold. Your stepping on my toes neither explains nor justifies my stepping on your toes unless I believe you stepped on my toes, *but the belief alone, true or false, explains my action*." (1963: 690) Given that *rationalizations* for Davidson only require invoking "primary" (motivating) reasons—which are belief-desire pairs for him—it seems he would also grant that an action can be rationalized even if the beliefs from which it proceeds are false. Clearly, though, he is suggesting in this quote that a belief can only *justify* an action if it is *true*.

(1) itself needs to be distinguished from correctly responding to all the *possessed* reasons. As I argue at the outset in Chapter 2, this, too, is distinct from rationality, though it is closer in some cases.

While it goes without saying that ethics and epistemology are substantively different enterprises, they are both normative enterprises. It is reasonable to expect a formal symmetry between them for this simple reason, and specifically with respect to the sorts of high-level notional distinctions that can be drawn. If drawing a distinction like this makes a big difference in ethics, it may in epistemology too. And if it doesn't make a difference in epistemology, that would call, I think, for explanation. While a few epistemologists have considered something *close* to this distinction, their treatment has not, as I'll argue in Chapter 2, been ideal.[3] The association of the suggestion with these particular theories may explain why most mainstream epistemologists avoid it. But it does not justify ignoring the conceptual distinction, which is a genuine one.

Drawing the distinction between (1) and (2) in the ethics and practical reason literature has led to a deep question. We are interested in how we really ought to act, and in what we really ought to intend. It is implausible to think that there are just distinct incommensurable practical oughts, none of which matters fundamentally.[4] Perhaps we could be forced into accepting this claim, but it strikes me as the last resort. What we really ought to do plausibly just is whatever we have most

---

[3]Foley (1987, 1993) recommended a distinction between the theory of epistemic rationality and the rest of the theory of knowledge, which he took to include the theory of whatever normative standing is necessary for knowing. But his theory of rationality is, we'll see, problematic. After Foley, the best discussion of rationality in the narrow sense in epistemology is found in Jackson (2011). Jackson is clearly aware of the broader literature that is my starting point, and uses it to great effect in debunking gappy arguments for "seemings internalism". His paper influenced my thinking in ways that will be obvious in Chapters 2 and 3.

      A *few* would say that in giving a theory of epistemic justification, they do not intend be giving a theory of epistemic rationality; see, e.g., Goldman (1986: 27). But the common tendency, particularly visible among internalists, has been to presuppose that justification and rationality trivially amount to the same thing. Huemer (2001: 22) affords one example: "Another word for what is justified […] is 'rational'." Cohen affords another in his formulation of new evil demon problem: "'Reasonable' and 'rational' are virtual synonyms for 'justified' […]. [C]learly the important concept, the one epistemologists have been concerned with, is […] 'rationality'." (Cohen (1984: 283–4)) BonJour affords yet another in his classic discussion of clairvoyance: "Samantha is being thoroughly irrational and irresponsible in regarding the evidence that the President is not in New York City on the basis of a clairvoyant power […] and this irrationality is not somehow canceled by the fact that she happens to be right. Thus, I submit, Samantha's irrationality and irresponsibility prevent her belief from being epistemically justified." (BonJour (1985: 39)) Another striking case is Fumerton (2006)'s introductory book, which contains a chapter entitled "Epistemic Rationality and its Structure" that has almost no further occurrences of 'rationality' but many of 'justification'. It is, I'll argue, no surprise that internalists like BonJour, Cohen, Fumerton and Huemer slide between talk of rationality and talk of justification. They cannot easily live without this slide.

[4]Some appear to say otherwise. I often hear ethicists say that there are different roles to be played by *subjective* and *objective* oughts, and that both are "important". They may justify their claim by saying that the subjective 'ought' goes with *blameworthiness,* and the objective 'ought' with *liability*. Well, if stipulatively fixing the reference of these terms *via* clearly distinct conceptual roles is all they intend to do, I can't complain. What I don't understand is the thought that neither of the oughts is more fundamental. Moreover, I see no point in stipulating distinct oughts. Just *directly* use the conceptual roles that fix the references and talk about non-motivational blame/praise-determining factors rather than subjective oughts. This is simpler.

      Some may quip that there are linguistic intuitions that speak in favor of positing two oughts. They say: "In Williams's case, there's a sense in which Bernie ought not to take a sip, and a sense in which he may, and that's that." Well, I agree that Bernie ought *rationally* not to take a sip. If that is the "sense", fine. This does not solve any problem. The question then is whether rationality is *robustly* deontic. Note that we can get lots of other "senses" with adverbial modification—'ought *legally*', 'ought *conventionally*'—and can in some contexts use 'ought' to pick out just these. Obviously, it does not suffice to show that conventional rules are robustly deontic that we can (in the right context) use 'ought' to mean 'ought conventionally'.

      Others think subjective oughts are needed to understand "action-guidance". Now, it is doubtless true that in acting for reasons, subjective factors have to enter into the picture *somewhere*. One cannot act *for* a reason—i.e., be motivated by it—if one hasn't the faintest clue that it exists. But this hardly shows that *when* one is acquainted with the reason, it can't guide action by being the *content* of a mental state. In cases where the content of one's mental state is not a good reason, one is not guided by a *new species of normative fact*—viz., the subjective species. One is rather guided by what *merely appears* to be a normative fact.

reason to do. If reasons and rationality can come apart in the way intuitively suggested by Bernie's case—and even more radically in other cases, as I note in Chapters 2 and 3—it becomes unclear why it matters whether we act rationally. In particular cases, this may not help us in the slightest to do or intend what we really ought to do or intend. Sure: we can often be *excused* in these bad cases. So it may not matter for our *reputation*. But that isn't what really matters.

Deeper developments of this simple thought have led some major figures to doubt whether we really ought, even *prima facie*, to be rational *per se*. This is a surprising conclusion. As Niko Kolodny—a central skeptic about the significance of rationality—notes, we seem not merely to be speaking in the register of *appraisal* but in the register of *advice* in using talk of rationality and irrationality. We don't seem just to be praising or blaming each other, but to be saying something with guiding significance. Intuitively, rationality is—to use some terms familiar to ethicists that I'll explain in Chapter 1—not merely **evaluative** or **hypological**, but **deontic**. And it seems *robustly* deontic. In noting that you violated a requirement of rationality, we seem to be doing more than noting that you've failed to comply with a mere standard of correctness, like of etiquette or grammar. Given the distinction between acting/intending for good reasons and acting/intending rationally, and the connection between 'ought' and '(has) most reason', the problem of explaining how this could be so is what some call the "normativity problem". I call it the **problem of deontic significance**, since I prefer to use 'normative' broadly. The problem isn't, as I'll stress in Chapter 3, a problem about the normativity *in the broadest sense* of rationality.

As it happens, I suspect there is no general solution to this problem about practical rationality. Arguing for *that* claim will not be my goal, though at times I'll gesture at why I think it is true. My goal is instead to argue that there is a *parallel problem* about epistemic rationality that epistemologists must address, and to present a solution to it that has broad implications for many epistemological disputes. In my view, there turns out to be *substantive* disunity between ethics and epistemology on this score: while epistemologists need to acknowledge that there is a *notional* difference between what a person would be rational or irrational in believing and what there is most reason for someone to believe or disbelieve, the extensions of these concepts are *far closer* than the extensions of the corresponding concepts in the practical sphere.

But I think it takes a *lot* of foundational work to see why this is true. Its truth should be viewed as both extremely surprising and extremely important for epistemology. Hence my project.

2.   *Problems about the deontic significance of epistemic rationality*

What is the problem about the deontic significance of epistemic rationality? Well, I ultimately think there are *several* problems, some more general and fundamental than others. I'll merely sketch one of them in this Introduction, saving a full survey of the problems for Chapter 3. (Some rest on more controversial assumptions that I'll motivate in Chapter 1, such as the assumption that epistemic reasons can be extra-mental facts.)

One problem—perhaps the core problem—about the deontic significance of epistemic rationality is related to what I find the deepest worry about internalist theories of epistemic justification. Internalists, recall, typically accept at least one of the following two views:

> *Supervenience Internalism*: Necessarily, for any subjects S and S*, if S and S* are nonfactive mental duplicates, S and S* are duplicates *vis-à-vis* what they are justified in believing.

> *Access Internalism*: Necessarily, for any subject S and condition C, the obtaining of C affects what S is justified in believing *only if* S could tell whether X obtains by reflecting solely on her nonfactive mental states.

I think one would get a truth if one replaced "justified in believing" with "would be rational in believing" in *Supervenience Internalism*, and that one *might* get one by doing so for *Access Internalism* too. Epistemic rationality is internalist in some major senses of "internalist", as I will argue in Chapter 2. But this is no reason for internalists of the familiar stripe to be excited. For there is a puzzle a lot like the one in the philosophy of practical reason about seeing why it *matters* from the epistemic point of view whether we are epistemically rational.

Being epistemically rational does not necessarily do anything to advance what epistemologists standardly regard as the **goals of cognition**, or to "make things go better" from the epistemic point of view. Indeed, rationality can, if it satisfies *Supervenience* and *Access Internalism*, **systematically** require adopting unreliable methods and abandoning reliable methods. In seeing this kind of point, one can turn Cohen's "new evil demon" and BonJour's clairvoyance objections to reliabilism on their heads. If a subject whose cognitive processes are massively unreliable can count as perfectly rational, and a subject whose cognitive processes are perfectly reliable can count as irrational, then *how,* given that our interest as cognizers would seem to be *hitting on the truth*, could it be epistemically *right* for us to aspire to live up to the standards of rationality as such? How could it be right to systematically "make things go worse" from the epistemic point of view and frustrate the "goals of cognition" *as a matter of principle* in some cases? The problem isn't, as I'll argue in Chapter 3, merely one that arises for a view on which true belief is the only final epistemic value, though plenty of internalists accept this view. There is a problem even if knowledge and understanding are final epistemic values. Rationality can force one to give up or shun knowledge-conducive and understanding-conducive methods that could otherwise be kept or acquired. (Indeed, I think more controversially that rationality isn't always required for knowledge or for understanding. If not, rationality cannot be necessarily desirable as a *constitutive* rather than *instrumental* means to these important epistemic standings.)

I think this puzzle is very much worth taking seriously, and that it transcends the interests that people have typically had in internalism/externalism disputes.

Typically the aim of these disputes was **extensional adequacy**, *modulo* typical theoretical virtues like simplicity and non-arbitrariness. Just get the intuitions about the application of 'rational' or 'justified' right, and you've got an important view, at least if it's not *ad hoc* or gratuitously complicated. Internalists can, I believe, get the extension of 'rational' right in a non-arbitrary, simple way. But this is *not enough*. Internalists must also show why a term with this extension picks out a property that has genuine deontic significance. They shouldn't, I think, be satisfied with an *irenic dissolution* of traditional internalism/externalism disputes—one that would, say, amount to claiming that internalism is right about the extension of 'rational' and externalism is right about the extension of 'justified', and leave it at that. The idea that the internalist's favored property is seriously deontic is crucial to the importance of her views for epistemology. Indeed, it is often used to *motivate* the theory. It would be a Pyrrhic victory to rest content with admitting that this property is not at all robustly deontic, but rather in a different category—say, a category of normative concepts associated with **praise** and **criticism** rather than **permissibility**.

So the internalist is left with a question much like the one that arose for those of us who accept that Bernie would be perfectly rational in taking a sip: even if his view is right, how could epistemic rationality have any deontic significance from the epistemic point of view? How could it matter from the epistemic point of view whether we avoid epistemic irrationality?

As I see it, this is the deeper puzzle behind more familiar disputes between some externalists—particularly reliabilist externalists—and internalists about epistemic justification. Justification, as I claim in the first chapter, is a deontic property that can be analyzed in the *ex ante* sense as *having*

sufficient (epistemic) normative reasons, and, in the *ex post* sense, as *believing for sufficient (epistemic) normative reasons.* Any satisfactory account of epistemic justification owes us a story about what it takes for reasons to be *good* reasons. Externalists have a clear account that seems obviously appealing, given a tempting picture of the goals of cognition and epistemic value: roughly, an epistemic reason is good when complying with it would make it *objectively more likely* that one would believe truly, and so *objectively less likely* that one would believe falsely. On some attractive externalist theories, complying with a sufficiently good reason for belief can get one very close to *knowing.*[5] Internalists, as I'll argue in Chapter 3, can have no similar story. Given the existence of a distinction between believing rationally and believing for good reasons, and the connection between good reasons and 'ought', it becomes tempting to draw the same kind of skeptical conclusion about the deontic significance of the property with which the internalists would have us identify epistemic justification—to conclude that it isn't epistemic *justification* at all, and that, while it may be rationality, rationality is **deontically impotent** *per se.*

3.   *The generality of the problems of deontic significance*

It is worth stressing that once one distinguishes between epistemic rationality and other epistemic standings, the problem at issue isn't merely one that confronts internalists. My own philosophical autobiography just inclines me to start the discussion in this way, since I once ditched internalism about epistemic justification precisely on the basis of supposed analogies with the problem for practical rationality. There are, I believe, problems about the deontic significance of epistemic rationality appreciable within otherwise attractive externalist theories.

One issue is closely related to worries some have about how such theories can *honestly* accommodate our intuitions about defeat. Suppose S comes to believe that P at *t* in virtue of a paradigmatically unconditionally reliable process in the presence of no defeaters. And suppose S wakes up on some later day having a strong new memory impression that the process by which she came to believe that P was unreliable. Suppose further that the impression has just haphazardly shown up in her mental life *via* a quirky subpersonal process rather than by genuine mnemonic processes. All the same, it has all the appearance of a real memory, and S can't tell otherwise. It is intuitive that S would be irrational in continuing to believe that P. Just *assuming* that this also influences *ultima facie* justification, some reliabilists add a clause to their theories intended to accommodate this intuition. They have it that this memory impression is an "undermining feature of S's cognitive state", which *ultima facie* justification does not tolerate.[6]
But there is a question about *how* this clause fits, once rationality and responsiveness to reasons are acknowledged to be distinct phenomena. How could rejecting a perfectly reliably formed belief on the basis of what, by the standards that motivated the theory in the first place, is a *terrible reason* be the right thing to do, epistemically speaking? Given that it is what rationality requires, this is a further way to see why we should worry about its genuine deontic significance.

This is one example of the broader question that comes to light when we start separating views about good reasons from views about rationality, as I think we ought. Of course, externalists are only generally committed in point of definition to a fairly weak claim: *not all* of the factors that affect justification—centrally, the factors that make for *good* reasons for belief—are internal, in some sense of 'internal' relativized to a particular thesis like *Supervenience Internalism* or *Access Internalism.* They are not definitionally barred from allowing that many purely internal factors— factors shared by nonfactive mental duplicates, or subjects who have reflective access to the same

---

[5]Bird (2007)'s view is like this: one is justified in believing P on this view iff one of one's nonfactive mental duplicates knows P. Insofar as Sosa (2007) intends to be giving an account of justifiedness as "adroitness", it is true of his virtue-theoretic proposal as well, since knowledge, for him, is accuracy *in virtue of* adroitness.
[6]See, for instance, Goldman (1986: 61–63; 109–113).

conditions—can influence justification. So, they face a question: *which* purely internal factors should be allowed to influence justification? If we are *not* skeptics about the deontic significance of rationality, we ought, I think, to allow many purely internal factors to influence justification. But if we *are* skeptics about the deontic significance of rationality, it becomes unclear why any purely internal factors should be allowed to influence justification *as such*.

So externalists need to decide whether to embrace skepticism about the deontic significance of epistemic rationality, and how far to go. One's theory of justification is not complete unless one has a theoretical rule for deciding whether, given a complete description of a subject's nonnormative circumstances and conditions (purely internal, historical and environmental), that subject is justified in believing that P, for any P. One cannot decide on a rule without deciding which purely internal factors matter. One could cling to the view that *although* justification and rationality are not *identical*, irrationality does diminish justifiedness, and hence allow that the internal factors that clearly influence irrationality can influence justifiedness in this indirect way. But it is not a good idea to do this without defending a view on the deontic import of rationality.

What I will be arguing in Chapter 3 is that the deeper theoretical basis for the most successful versions of externalism actually motivates skepticism about the deontic significance of rationality, and about the significance of purely internal factors in a much more general way than is usually explicitly claimed by these theories. Process reliabilists should recognize that their criterion for what it takes for an epistemic reason to be a *good* reason is that complying with that reason would be an instance of a reliable process. Indicator reliabilists should recognize that their criterion for what it takes for an epistemic reason to be a *good* reason is that it is objectively indicative of the facts. Virtue-theoretic externalists should recognize that their criterion for what it takes for an epistemic reason to be a *good* reason is that the ability to heed reasons of this kind is a competence, where competence is then understood (typically) in reliabilist terms.

This feature of their views stems from a deeper commitment to the seductive thought that the principles that determine whether reasons are good should be principles that help us to achieve our *epistemic goals*, which are often taken to be *decreasing false belief* and *increasing true belief*. No purely internal factors amount to good reasons *as such* given these criteria for goodness and this motivation for the criteria. Since the factors that affect rationality *per se* are, as I'll argue, purely internal, they ought to commit to a sharper division of rationality and justification, to skepticism about the deontic significance of rationality, and to the deontic irrelevance of the purely internal.

4.   *Teleology and the source of the problems in the epistemic domain*

Now, I find the conclusion that epistemic rationality is deontically impotent harder to swallow than the parallel conclusion about practical rationality. For a year or so I accepted this conclusion. But I became increasingly uneasy. While I ultimately embrace the skeptical conclusion about practical rationality, I don't think we are forced to accept the same conclusion about epistemic rationality. Epistemology and ethics are different on this score for *substantive* reasons. So, after Chapter 3, I'll be attacking the fundamental presuppositions on which the main arguments for skepticism about the genuine deontic significance of epistemic rationality must rest. The presuppositions that I want to attack are fundamental, widespread ones that have rarely been questioned or defended, perhaps because they just seem obvious. But if they seem obvious, it's only, I think, because of subtle confusions or unstable implicit analogies with ethics.

Before explaining what these assumptions are, I'll stress that I am happy to agree with reliabilists and many others that truth, understood as a property of doxastic representations, is a central intrinsic epistemic value. Indeed, I'll give a novel defense of the stronger view that it is the *only* intrinsic epistemic value in Chapter 4. What I want to argue is that this view and its implications

for the structure of deontic facts in epistemology have been misconceived. Failing to question the misconception must lead, I believe, to skepticism about the deontic significance of epistemic rationality. We can avoid this skepticism only by rethinking the place of truth in epistemology.

As an epistemic value, truth—understood as property of doxastic representations—has almost always been understood **teleologically**. It is often presupposed that *qua* epistemic agents, we ought to value it just as a goal or end. *What it is* for truth to be epistemically valuable on this view is for it to be something we ought epistemically to increase in our doxastic inventories, and *what it is* for falsity to be epistemically disvaluable, on this view, is for it to be something we ought epistemically to *eliminate* from our doxastic inventories. This is a "veritist" version of what I'll call **weak epistemic teleology**. Weak epistemic teleology itself is broader view about what it is for something to be epistemically (dis)valuable. One sees it in conceptual slides like these ones from Alston (2005: 29):

> We evaluate something epistemically when we judge it to be more or less good or bad from the epistemic point of view, that is, *for the attainment of epistemic purposes*.

> [T]he evaluative aspect of epistemology involves an attempt to identify ways in which the conduct and products of our cognitive activities can be better or worse *vis-à-vis the goals of cognition*.
>
> (Italics mine.)

Weak teleology is contained in the shift from talk of *good* and *bad* to the talk of *purposes* and *goals*. This shift is certainly not local to Alston. One sees it all over the place in epistemology.

This is not the only kind of teleology that has been dominant. Besides assuming that epistemic *value* is teleological, many epistemologists assume that epistemic *norms* are, at least indirectly. On this view, the way we judge whether a rule of belief-formation is *right* is by considering whether following the rule would be *in the service of epistemic value*: whether it would, for instance, lead in the long run to more true beliefs and fewer false beliefs, if these are the chosen values/disvalues. This assumption I call **strong epistemic teleology**. It may also be called *epistemic consequentialism* in a broad sense, if we follow the convention of calling "consequentialist" normative theories that make the *good* explanatorily prior to the *right*, and that make facts about the right supervene on facts about the good and clearly nonnormative facts (e.g., about means-end relations).

Properly understood, I think both kinds of teleology have no *fundamental* place in theorizing about epistemic normativity and value. Indeed, I think these views get things almost exactly backwards. This will be for reasons that, as far as I know, have never been appreciated; epistemic teleology has, of course, been recently attacked by Selim Berker, but I am unconvinced by his objections for highly general reasons I'll detail in Chapter 4, where I will recommend a completely different strategy. The center of this book—Chapters 4, 5 and 6—will be an attack on both kinds of epistemic teleology, and a development of an alternative picture of the epistemic value of truth that is modeled after themes in nonconsequentialist ethics. It in effect analyzes epistemic goodness in terms of epistemic rightness rather than *vice versa*, and understands both in nonteleological terms. As it happens, this is one reason why I think there are substantive disanalogies between ethics and epistemology. I'm seduced by broadly teleological views about many practical values and norms. Together with a mistaken tendency to assume that there will be substantive rather than merely formal analogies between ethics and epistemology, the seductiveness of such views—the feeling that they state *platitudes* about the relation between value concepts and deontic concepts—may be what leads epistemologists to presuppose both kinds of teleology in thinking about epistemic value and the relation between it and the structure of epistemically deontic facts. As I'll show, rejecting these bedrock views in the way I suggest is the key for seeing how epistemic rationality could be genuinely deontic.

5.   *Non-teleology in the service of epistemic rationality*: *the limits of goal-driven epistemology*

Now, you might ask: "If our fundamental epistemically normative relation to truth is not a teleological one, what else could it possibly be?" I think it is a simple one: truth commands **respect** as a standard of believing in the particular case. More specifically, my view will be that the only fundamental requirement we are under *qua* epistemic agents—the only basic one compliance with which is *constitutively epistemically* required—is *to respect the rule to believe P only if P is true*. On this view, the story about why epistemic rationality is seriously deontic is straightforward: complying with requirements of epistemic rationality *just is* a large constitutive part of governing our doxastic deliberations in ways that respect this standard.

On my view, the epistemic value of truth, understood as a property of doxastic representations, is spelled out fully by the requirement to respect the rule to believe that P only if P. To claim that such a requirement exists is all that can reasonably be involved in saying that truth is an epistemic value, at least if this is supposed to be a claim that has any direct implications for normative epistemology. My view therefore predicts—plausibly, I think—that it is not bad from the epistemic point of view if one forms no true beliefs simply because one forms no beliefs. If you are not in the business of belief-formation, you can't fail to respect truth as a standard, and so can't fail to correctly respond to epistemic value. There is no *epistemic* obligation to form any beliefs whatsoever, or to be interested in any questions whatsoever, or anything of the kind.

Respect for truth may also not always forbid forming a false belief. This could hold, for instance, if sufficiently good but misleading clear evidence recommends believing something that turns out to be false. So my view also does not *entail* that false belief is *always* bad even *pro tanto* from the epistemic point of view. This may seem less attractive. But the failure of my view to entail that believing falsely is *always* bad from the epistemic point of view is no objection if we bear in mind an ambiguity in "bad from the epistemic point of view" I'll discuss in Chapter 4.

False beliefs are often bad *period*. Being epistemic items in a broad sense, they might be called "epistemic bads" in a *purely classificatory sense*—i.e., bad and, *by the way*, epistemic. If we have an independent non-instrumental concern about such epistemic items—say, because we want knowledge for its own sake as *a good* of the sort that might figure in some Moorean list including *pleasure* and *beauty*—we might even say that they are "bad from the epistemic point of view", again in a purely classificatory sense. Bad, since they prevent us from knowledgeably answering certain questions, where the answer is of intrinsic interest. The crucial point is that *this* is not a kind of badness that is relevant to the correctness or incorrectness of epistemic norms.

Indeed, I'll argue that the conflation of these two senses of "bad from the epistemic point of view" is at the heart of the attraction to teleological pictures of epistemic values and norms. For there is a *different* sense—one that is irrelevant to normative epistemology though certainly not to practical life—in which true (or false) belief might be valuable (or disvaluable) in a teleological way. So, seeing why the worry is not serious will be key to developing an *error theory* about how it is that teleology came to be so dominant in epistemology. There certainly is a sense in which true belief might be a teleological value. But it is also not a sense that has any relevance for epistemology properly conceived as a normative enterprise with its own standing.

This distinction will be part of a broader theme that motivates my particular view, and my overarching opposition to teleology. It should, I believe, be striking that evaluative concepts in the narrow sense—concepts like *value* and *disvalue*—have not had a proprietary place in epistemological discussion until *very recently*. Historically, the focus in epistemology was dominantly either on *deontic* notions—e.g., epistemic *justification* and *ought* and *reasons*—or on particular epistemic standings—e.g., *knowing*. It should also be striking that the recent literature on

"epistemic value" grew out of a renewed interest in the comparative value *simpliciter* of various epistemic standings. It grew out of the question raised by the Meno Problem—viz., why we value knowledge more than mere true belief if mere true belief would get us just as far on the road to Larissa. If epistemic value is understood as something to be studied by *epistemologists in particular* rather than general axiologists, that problem cannot really be about epistemic value, as I argue (along with Michael Ridge), and as should be obvious from what I already said.

I think there is a simple explanation for the long absence of proper value concepts in epistemology: epistemic value and disvalue, understood in the only way that has a proprietary location in epistemology, could only be highly derivative. There is an analogy here with "thinly" normative domains. Consider etiquette. This is first and foremost a system of *rules*. It purports to tell us what we *conventionally ought to do*, or "what is *done* and *not done*". It is not primarily concerned with *value* except in a sense that has to be *stipulatively defined* in terms of the conventional 'ought'. "Good behavior" from the point of view of etiquette could only mean *behavior in accord with the rules*. We can *choose* to talk about what is "good from the point of view of etiquette", or about what "conventionally good". But the way to spell out what that means is not by *discovering* an underlying axiology. It is by *erecting* an axiology on the deontic foundations.

Epistemology, I think, is structurally analogous. Insofar as it doesn't just mean "value *simpliciter* that happens to be epistemic", epistemic value has to be understood by reference to distinct epistemically normative concepts. True belief is an "epistemic value" in the relevant *proprietary* sense only because truth happens to be the standard for belief that we ought epistemically to respect. If the norm of belief at issue is not a teleological one—as I think it isn't, like other norms of respect—it is no surprise that weak epistemic teleology must fail, and that strong epistemic teleology gets things backwards. There is no sense in trying to found epistemically deontic facts on an epistemic axiology. The latter is in large part to be *derived*, not *discovered*.

This is another respect in which I think there is substantive disunity between epistemology and ethics. Ethically informed readers may hear Philippa Foot's voice in the last paragraph. I borrow the analogy with etiquette and the dialectical posturing from her.[7] As I've already indicated, I am *not* opposed to teleology in ethics. The concept of welfare is an evaluative one that is perfectly intelligible on its own terms—and on teleological ones. And it can be used in formulating a *prima facie* standard of correctness for ethical rules—though it is perhaps not the whole story, as some utilitarians would insist. Still, while I think Foot was wrong in this debate (which is not something I'll try to defend), that would be a substantive point about ethics. We should always bear in mind the possibility of something structurally like her suggestion validly applying to *other* domains. It is, after all, the right thing to say about "good from the point of view of etiquette". It is also, or so I'll be arguing, very close to what we should say about "good from the epistemic point of view" in the relevant sense—the proprietary sense that puts epistemic value into the hands of

---

[7] Foot sometimes makes stronger claims than the one I made in the last paragraph. In Foot (2002: 101), she seems to think the notion of what is good from the point of etiquette is *unintelligible:*

> It will be helpful here to consider a parallel between the case of morality and the case of etiquette. Why is it […] that we do not have good and bad states of affairs from the point of view of etiquette? The notion seems incomprehensible. But why is it so? It is so because a good state of affairs from the point of view of etiquette would be one which from the point of view of etiquette it is right to aim at or produce. But in fact there can be no state of affairs that stands in this position because even if there are aims prescribed by etiquette […] etiquette is also a matter of following the rules, and the rules circumscribe the way in which the aims may be followed.

This goes too far. It is better to say that rules come first in etiquette, and that states of affairs will be better or worse from the point of view of etiquette to the extent that they are in harmony with these rules. It is not that "good from the point of view of etiquette" is unintelligible, but that it is essentially *secondary*.

epistemologists rather than of general axiologists.

If I am right about this, it would show the limitations of a robustly "value-driven" epistemology. I do agree that our intuitions about the value *simpliciter* of epistemic standings like knowledge, particularly in comparison with true belief, can impose constraints on theories of knowledge. If it is true that knowledge is better *simpliciter* than true belief, a theory of knowledge that lacked the resources to explain this would be *ipso facto* worse off. But it is *another* thing to think that value could play the kind of role in normative epistemology that it has often been taken to play in normative ethics. To play a structurally analogous role, we would need to be talking about *properly epistemic* value, not just value *simpliciter* that happens to be epistemic—even final value *simpliciter* that happens to be epistemic. Moreover, properly epistemic value would have to be understandable in properly epistemically teleological ways for it to play the kind of role that value *simpliciter* plays in consequentialist ethical theories. It is this that I think cannot be done. If that is true, the way in which epistemology could be "value-driven" is going to be much more circumscribed than the way in which ethics plausibly may plausibly be "value-driven".

6.   *An overview of the following chapters*

Having summarized some of the themes, goals and views that I will be developing, here is a breakdown of how I will proceed.

In Chapter 1, I'll lay down some distinctions and assumptions about normativity that will be needed throughout the book. I explain the division of normative domains into what I call, following ethicists, *deontic*, *evaluative* and *hypological* categories. I express sympathy for a view that explains why all of these categories count as normative ones—viz., *Reasons Basicness*, which holds that all normative facts are reducible to facts involving reasons and nonnormative facts. I'll defend a high-level account of epistemic justification that appeals only to epistemic reasons and nonnormative facts. Having this view ready-to-hand makes it easier for me to erect useful structural parallels between epistemic normativity and practical normativity that bring to light the problem about the deontic significance of rationality that I want to address in later chapters. I will also set the stage for the discussion of epistemic teleology and my "epistemic Kantian" alternative to it in later chapters. I'll conclude with a brief discussion of hypological concepts (i.e., concepts associated with the criticism and praise of people for hosting certain attitudes or doing certain acts) and their distinctness from deontic concepts. This also matters for later purposes, since, as I suggest in Chapter 2, the concept of rationality is a hypological concept in the first instance, and skeptics about the deontic significance of rational requirements would do well to take advantage of this fact.

In Chapter 2, I rehearse why philosophers of practical reason have been led to separate a person's practical rationality from her responsiveness to normative practical reasons. I do this in order to set the stage for a similar separation of epistemic rationality and responsiveness to normative epistemic reasons. Along the way, I object to the purely belief-relative view of practical rationality that many people in this literature have come to accept, and erect a dilemma to show that all purely belief-relative views fail. But I note that the core motivating thought behind these views— that what it is rational to do and intend turns on *apparent reasons*—can be captured in a different way, if we take appearances to be nondoxastic seemings.

On the replacement view, it would be rational for someone (to intend) to A *only if* the practical reasons that would exist if the person's total set of nondoxastic appearances were veridical would sufficiently support A-ing. This is only a *necessary* condition for practical rationality: full practical rationality requires not only responsiveness to apparent reasons in this sense, but also satisfaction of (wide-scope) requirements of coherence. I recommend a parallel view about epistemic

rationality on which it would be rational for someone to believe that P only if the epistemic reasons that would exist if the person's total set of nondoxastic appearances were veridical would make it objectively likely that P is true. I explain why this view is preferable to other accounts, and how epistemic rationality could easily come apart from epistemic justification if this view and the high-level account of justification from Chapter 1 are true.

In Chapter 3, I rehearse why philosophers of practical reason have been led to doubt that practical rationality has genuine deontic significance. While their explicit arguments for this turned nominally on their acceptance of the purely belief-relative view of practical rationality I attacked in Chapter 2, I explain why all the same the doubts remain if one switches to the appearance-relative view I recommend. I then explain how these doubts can be taken to generalize to epistemic rationality. But I note that the generalizability of the doubts turns heavily on the truth of what I call *strong epistemic teleology*, and hypothesize that the doubts wouldn't arise if one accepted my alternative view. Indeed, I think the doubts definitely do not arise if one accepts that view, but I save the full argument for that claim for Chapters 4 and 5.

I turn in Chapter 4 to argue against strong epistemic teleology. I argue against it by arguing against a weaker view that it implies—the view on which epistemic value, and in particular *truth*, understood as a property of doxastic representations, is to be understood as something "to be promoted" and pursued as a goal. In the course of doing this, I argue at more length that truth is not correctly viewed an epistemic goal at all. Indeed, I think there are *no* properly epistemic goals in any sense that could be relevant to normative epistemology. I argue that the idea that belief constitutively aims at truth does not motivate this claim when that idea is correctly formulated. Truth, I suggest, is a standard to be respected rather than a goal to be achieved. In the course of arguing against the weak teleology implied by strong teleology, I also consider direct arguments against the latter advanced by Selim Berker. I suggest that my strategy is better, since Berker fails to anticipate some extremely natural replies to his objections. Indeed, I think there is something fundamentally amiss with Berker's objections, and seeing what it is turns out to be crucial for answering an objection to my own epistemic Kantian view.

In Chapter 5, I explain in much greater detail how I understand the place of truth in epistemology, and develop and motivate my epistemic Kantianism in much greater detail. This leads well into Chapter 6, where I explain how accepting my picture of the place of truth in epistemology can solve the problems about the deontic significance of epistemic rationality. It also leads well into Chapter 7, where I explain how epistemic value theory becomes a secondary enterprise once the fundamental epistemic value is understood—as it should be—in non-teleological terms. I express strong skepticism about the very idea of "value-driven" epistemology in this chapter. And I develop at much greater length the claim suggested earlier in §5 of this introduction that evaluative categories have no fundamental role to play in epistemology, precisely because they need to be analyzed in deontic terms.

I then explore the implications of my view for some specific debates in epistemology in Chapter 8, including internalism vs. externalism, debates about how defeat ought to be understood within externalist theories, debates about the deontic significance of "higher-order evidence" and about the significance of disagreement. I conclude by sketching a vision of how epistemology as a whole ought to be redirected in the light of the earlier chapters.

I

REASONS AND OTHER NORMATIVE FACTS

While my ultimate focus will be epistemic *rationality* and some puzzles about its significance, I'm going to start by explaining how to understand normativity in general, though with the spotlight on the epistemic domain. On an increasingly dominant approach in the broader literature on normativity to which philosophers of practical reason more often contribute, *reasons* are seen as playing a bedrock role: all normative facts reduce to facts involving only reason-relations and nonnormative properties and relations.[8] I aim to extend this approach to epistemology in a way that is sensitive to worries I'd expect mainstream epistemologists to have. I'll defend a neutral, meta-normative account of epistemic justification that internalists and externalists alike can accept: one on which propositional justification amounts to *having sufficient epistemically normative reasons*, and on which doxastic justification amounts to *believing for* such reasons.[9] While I expect some to balk

---

[8]This, of course, leaves it open whether facts about reasons ultimately reduce to naturalistically acceptable facts. Our meta-normative claim just aims to greatly simplify the metaphysics of normativity. But it *might* facilitate further reduction. Mark Schroeder suggests as much in his (2005) and (2007: ch. 4).

[9]This account is *neutral* because it takes no stand on how epistemically normative facts supervene on the *nonnormative* facts, but *meta-normative* because it takes a stand on what the more fundamental normative facts are within the epistemic domain. It is just a view about *how certain normative facts reduce to others*.

Oddly, some epistemologists have often not seemed to see the *bare possibility* of this type of view.

at these claims, the central reasons for balking are, I'll claim, at bottom confused. Having defused the worries with attention to distinctions that are more familiar in the practical literature, I'll briefly explain how epistemic justification may relate to (and come apart from) *epistemic value* and *responsibility,* largely to set up a framework for later chapters.

This all matters for the big picture. As I say in the Introduction, my ultimate aim is to generalize to epistemology some problems about the significance of rationality that are familiar in the practical literature, to show how this leads to a reorientation of key disputes, and to try to solve these problems in a way that vindicates *some* internalist themes by undermining some rarely questioned but crucially false foundational assumptions. These problems about rationality from the practical literature have tended to turn on the view that reasons come first, and that other apparently normative properties—esp. rationality—are best analyzed *indirectly* in terms of reasons. To create symmetry between the epistemic and practical cases, it is helpful to extend the assumption that uncovered the original puzzles to epistemology. This is a real task, since reasons are not yet a "common currency" among epistemologists in the way they are among ethicists and philosophers of practical reason.[10] Hence this chapter.

1.   *Three faces of the normative: deontic, hypological, and evaluative*

There are several different types of normative fact. Following (most) people working in ethics and the philosophy of practical reason, I'd ask us to recognize at least three. Firstly, there are

> **deontic** facts, like the fact that A-ing is *obligatory*, that B-ing is *permissible*, that C-ing is *wrong*, that D-ing is *justified*, the fact that there are *sufficient reasons* to E*,* etc.,

where A−E are acts or attitudes. Certain species of deontic fact are intensely studied in ethics, where a long-standing goal has been to specify the most general moral and prudential deontic facts, which explain all the others. But they appear in other domains, including epistemology.[11]

Consider how evidentialists are sometimes criticized for having a circular or trivial account of justification. Is this because facts about what is evidence for what are *normative* facts? *That* objection would fail. It would badly overgeneralize. If it were any good, there could be no non-trivial question about metaphysical priority within the sphere of normative facts. There *is* such a question, and it is an extremely interesting one. It is often asked within the broader literature, which is admittedly written mostly by philosophers of practical reason. If the objection we're imagining worked, all these people would be engaging in a trivial or circular enterprise. This is not so. Evidentialism ought, as I'll argue later, to be a meta-normative theory. Facts about evidence *just are* facts about certain kinds of good reasons, and such facts are clearly normative. Analyzing justification in terms of evidence even with this explicit proviso remains *extremely non-trivial.*

So there is a simple answer to the circularity charge one sometimes sees. Bits of evidence are, yes, epistemic reasons of a certain kind. Evidentialism claims that facts about justification reduce to facts involving such reasons and some nonnormative facts. This is not close to being a trivial view, as we'll see.

[10]I steal the nice phrase "common currency" from Reisner and Steglich-Petersen (2011: xii), whose volume on reasons for belief might, I hope, be the beginning of a turn to reasons in epistemology.

[11]It is obvious that we often say things of the form 'You ought to believe that P' and 'You shouldn't believe that Q', and extremely intuitive that such claims are often true when read as involving *distinctively epistemic oughts*. Of course, some have worried that the truth of such readings of these claims would require the truth of the implausible doctrine of *doxastic voluntarism*. But these worries are, I believe, deeply confused and badly overgeneralize. I argue for this, at any rate, in Appendix A.

My take is essentially the same as that found in McHugh (forthcoming), and I'd advise any epistemologist who wants to think clearly about these issues to read his fantastic paper. A rather quick way of putting the point starts with the observation that everyone should agree it is often true that you ought to intend to A, ought to desire to B, ought to feel good about the prospect of X's occurring, and so on. Yet desires and feelings are not generally subject to voluntary control. The Toxin Puzzle suggests that intentions aren't either. So we *already* needed a different model for thinking about how oughts apply to *attitudes.* The right analogue for thinking about how oughts apply to beliefs is *not* found by focusing on how oughts apply

One aspect of all such facts is worth noting immediately. Deontic features apply most directly to *acts* and *attitudes*. It's things like *killing* or *intending to kill* that are wrong or prohibited, and things like *helping* or *intending to help* that are required or right or justified. Of course, these are required *of* people, permitted or justified *for people*, etc. But these people aren't the fundamental targets of deontic evaluation. So I'll call deontic facts (**primarily**) **act/attitude-oriented**.

Deontic facts here contrast with a second class of normative facts. These facts turn on the evaluation of people in doing acts or having attitudes. I'll say that these facts are (**primarily**) **person-oriented**. Such facts are not often given a proprietary name. Michael Zimmerman calls them **hypological**. Although not many have picked up on his terminology, it is worth having. I'll use it, mainly for lack of any helpful alternative.[12] Hypological facts include facts like

> the fact that *S is criticizable for A-ing*, that *S* is creditworthy for B-ing*, that *S** is excusable for C-ing*, and so on,

where S, S* and S** are people, and A, B and C are acts or attitudes. *How* hypological facts relate to deontic facts is an interesting question. They *uncontroversially* come apart as follows: an act/attitude can be wrong, unjustified, or insufficiently reason-supported while one is not criticizable for it ("*blameless violation*"), and an act/attitude can be right, required, or supported by decisive reasons while one is not creditworthy for it (call this, by symmetry, *praiseless nonviolation*).

Whether deontic and hypological facts can separate in even deeper ways is something I'll return to later. The uncontroversial gaps can crop up for at least two reasons. Firstly, it is plausible that S must in the first instance be A-ing *autonomously* for *her* to be open to evaluation in A-ing.[13] But irrespective of her autonomy, her A-ing can still be deontically assessed. Besides autonomy, ethicists also acknowledge some cognitive conditions for the applicability of hypological concepts to acts and motivational attitudes. If S acts wrongly but isn't in a position to appreciate this, many would take that to be a good excuse, at least in the first instance. If S acts rightly but S's evidence indicates that S is doing wrong, that makes S an inappropriate target of praise.[14]

---

to *actions,* but rather by focusing on how they apply to *other attitudes*. If there is a story to be told about how oughts can apply to other attitudes, we can apply the model to belief.

Then, the second prong: it is an altogether familiar point in post-Frankfurt action theory that freedom of *action* and freedom of *will*—of intention, of desire—are *different things*; cf. Frankfurt (1971) for the original observation. Oughts that apply to attitudes may presuppose freedom *of these attitudes*. But as Frankfurt has insightfully argued over his career, the model for *that* is different from the model for freedom *of action*. Precisely what form the model should take is hotly disputed in contemporary action theory. But here we have an *embarrassment of riches* if we are looking for ways of thinking about freedom of belief.

Old worries about doxastic voluntarism stem from an unfortunate ignorance of this broader literature and the distinctions it has uncovered. An argument against the possibility of oughts applying to emotions, desires, and intentions from the fact that we often lack voluntary control over them would be rapidly rejected in this literature. And for good reasons. The reasons for rejecting the Alston/Plantinga argument for skepticism about epistemically deontic facts are going—*surprise, surprise*!—to look very similar.

[12]Zimmerman (2002: 554). The only others I've seen use the term are Clayton Littlejohn and Ishtiyaque Haji.

[13]I say "in the first instance" for the following reason. Suppose you make yourself temporarily insane because you know that when you become insane, you will be able to cause a certain bad outcome you couldn't otherwise bring yourself to cause. Here your blameworthiness for the earlier intention plausibly *transmits,* so that you are still responsible for the outcome. This has come to be called *derivative* or *secondary* responsibility. (Cf. Rosen (2004) and Smith (1983).) But in the *primary* case, lack of autonomy in A-ing is a good excuse.

[14]At least on the most plausible view. For a remarkably long time in the ethics literature, people ignored the relevance of evidence to blameworthiness and praiseworthiness, and focused only on beliefs and motives. Sher (2009) is the first book *ever* written *just* on epistemic constraints on blameworthiness/praiseworthiness. Cf. his introduction for confirmation of the unfortunate state of the previous literature.

A question that I'll frequently address is how the idea of deontic-hypological separability extends to—and matters for—epistemology. Remarkably, the extension was barely even considered until fairly recently.[15] Epistemologists—particularly internalists—have often conflated the deontic and the hypological. While there are surely *substantive* disanalogies between ethics and epistemology, the mere high-level idea that these properties might be separable should not be seen as radical. It is a view we would expect if there is any *structural* unity between normative domains. The burden of proof is against blurring the deontic and hypological in epistemology. Ironically, I shall be vindicating the blur to some extent in later chapters. But I take this to be one of the most surprising conclusions I will push, and an aim in the initial chapters shall be to emphasize exactly how deeply surprising this conclusion ought to appear.

A final class of normative facts is the class of

> **evaluative** facts, such as the fact that *A is good* or *valuable*, the fact that *B is bad*, the fact that *C is desirable*, the fact that *D is sexy* or *kind*, the fact that *E is ugly* or *cruel,* etc.

Evaluative facts split into subcategories. They differ in their *thickness*, as Williams (1985) influentially put it. Thick evaluative properties are ones like being *kind* or *cruel, sexy* or *ugly*. They are thick because they seem to be layered, including some nonnormative and some purely evaluative aspects. Kindness, for instance, is linked with nonnormative features—patterns of motivation and behavior—which are viewed through an evaluative lens. In contrast, properties like badness and goodness are purely evaluative. All we have is the lens. So they are called *thin*.

Evaluative properties can also be subdivided to track a syntactical distinction whose importance for normative theorizing was stressed by Geach (1956). We can use 'good' either as an *attributive* or as a *predicative* adjective. We can say "This is a good knife", and we can also say "That state of affairs is good". This distinction in language seems to track a difference in normative reality. When we say "This is a good knife", it is not hard to see how we are making a *clearly naturalistically acceptable* claim. Knives have a function: to cut. A knife will be good to the extent that it can fulfill this function, bad to the extent that it can't. The "normative" facts come with the territory: they flow from the nature of the thing *qua* knife. Whether all attributive evaluative facts are grounded in the natures of things *in the same way*—as a matter of *function*—is a hard question. But the prospects for naturalistic reduction look pretty good regardless.

We cannot clearly say anything similar about the fact that some state of affairs is good. States of affairs are just, well, states of affairs. They have no point. People like Geach take this to suggest that claims about the goodness and badness *simpliciter* of states of affairs are meaningless or false. Others—e.g., Thomson (1997, 2008)—look at particular states of affairs, find particular attributive evaluative truths around them that make the general predicative claims true, and reduce all evaluation to attributive evaluation. Others—e.g., Scanlon (2011)—take the opposite strategy and try to ground attributive value facts in more substantive normative facts, like facts about normative reasons. Still others—e.g., Parfit (2011)—will say that neither reduces to the other. These same people will probably add that attributive uses of 'good' aren't normative at all: they are just *disguised descriptive claims*. The distinction will be important when we turn much later to discuss epistemic value and its relation to the deontic. A lot of the new literature on epistemic value has proceeded under confusions generated by a failure to draw this distinction.[16]

---

[15]Some of the first people to make a big deal about it in epistemology were Bach (1985) and Engel (1992); cf. Dancy (1992) for a nuanced early discussion that casts doubt on Bach and Engel's thought that internalism/externalism disputes partly trade on a conflation of the deontic and the hypological. Littlejohn (2009) has revived this thought. And the frequent invocation of the category of excusable violation in the post-Williamson literature on norms of assertion made it much more widely appreciated.

[16]See Ridge (forthcoming) for an excellent development of this accusation.

Besides the questions that we can ask *within* each of the three classes about which of its members is the most fundamental, we can also ask an important question about all the classes: what makes them all normative? In the last fifty years or so, there has been a trend in the practical literature to single out *reasons* as providing the answer. Raz put this bold answer nicely: "The normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons."[17] Following Schroeder (2007), I call the natural reading of this claim that entails that all normative facts reduce fully to facts about reason-relations and nonnormative facts **Reasons Basicness**.[18]

Reasons Basicness is obviously attractive for deontic facts. We can understand the differences between requirements, oughts, justifications and permissions as differences in the *strength of underlying reasons*: requirements consist in strongly decisive reasons, oughts consist in decisive reasons, justifications consist in strongly sufficient reasons, permissions consist in sufficient reasons, etc. The appeal for evaluative properties is less obvious. One way to turn the trick is with Scanlon's **buck-passing account of value**: goodness(/badness) is viewed as the higher-order property of having lower-order properties that provide reasons for pro(/con)-responses. A puzzle for this kind of view is how to sort out the reasons for responses that are intuitively value-grounding from those that aren't. This is the "Wrong Kind of Reasons" problem. I think this problem is soluble—or, more precisely, *dissoluble*—for reasons I discuss elsewhere, and so I think buck-passing is defensible.[19] Successfully applying Reasons Basicness to hypological properties is also conditional on a solution to this problem. But it is otherwise easy to see why one would be attracted to this extension. Hypological terms wear the relevant responses on their sleeves (e.g., "*blame*worthiness", "*praise*worthiness", "*criticiz*ability", "*apprais*ability", etc.). Why not suppose that these properties just are formal properties of having lower-order features that give reasons for these responses (e.g., having the reactive attitudes involved in blame and praise)?

Reasons Basicness has not yet taken over in epistemology.[20] This is an extremely surprising fact, I think, and it calls for some explanation. I will later offer some conjectures about why epistemologists might be avoiding it. While none of my later arguments will turn *critically* on Reasons Basicness, it will be *easier* to state them with this view in place. And so I hope to show how reasons could be given a central role in epistemology in this chapter and throughout.

Having sketched the core distinctions in a way that crosscuts domains, I'll now turn to discuss normativity in the epistemic domain specifically. My reason for starting in a domain-neutral way is that several of these distinctions have only clearly been made by people in the practical literature. The distinctions are also of great importance in epistemology. Many of my points will turn on the possibility of drawing them. If I had allowed myself to talk only in ways in which epistemologists have already tended to talk about normativity, I would have lacked the resources.

---

[17]Raz (2002: 67). In spite of this nice quote, it is unclear whether Raz himself accepts the claim that all normative facts *reduce to* facts involving only reasons and nonnormative facts. See Heuer (2004) for a subtle discussion of Raz's view that indicates that he may back off from this reading of the quote, and take facts about value to be as primitive within the normative domain as facts about reason-relations.

[18]See Schroeder (2007: 81). He nicely emphasizes how strikingly widespread the idea is: "[I] think that [Reasons Basicness] is a broadly attractive thesis about what is distinctive of the normative, something like which is currently accepted or at least found attractive by moral philosophers of extremely different bents. Any view that looks attractive to Jonathan Dancy, Michael Smith, Derek Parfit, T. M. Scanlon, and Joseph Raz has to be one with some kind of very broad appeal."

[19]See Sylvan (ms3).

[20]But we may be in for a change on this score. Skorupski (2011) takes Reasons Basicness to the limit and applies it to epistemology, as well as many places where one wouldn't have expected reasons to be relevant (e.g., modality (!)). Andrew Reisner and Asbjorn Steglich-Petersen's new collection <u>Reasons for Belief</u> is a self-conscious effort to make mainstream epistemology acquire a focus on reasons. In their introduction to the collection they are explicitly optimistic about extending Reasons Basicness to epistemology.

2.   *Reasons: some basic distinctions*

Reasons Basicness will be helpful for my overall project, even if it is not strictly required. I independently think that reasons afford a great tool for illuminating what makes normative facts normative. So I'm going to start with epistemic reasons, and I'll then turn to show how some other normative facts in epistemology ultimately boil down to facts about reasons.

Some points of clarification are in order first. We need to distinguish three possible senses of 'reason', only two of which will matter. Firstly, we can talk about the **reasons why** someone believes something or, in general, does something. It may be that the *ultimate* reason *why* you believe Jonathan Schaffer's monist view is that it sounds cool. The cool sound of the theory led you to be interested in seeking arguments for it. While these later discovered arguments provide the only reasons *for which* you believe the theory, the ultimate explanation of *why* you believe it would advert to how it sounded cool. We can call such 'reasons-why' **explanatory reasons**.

There is an important *proper subset* of explanatory reasons. Even if you were fully self-conscious and honest with yourself, you would not say that *your reason* for believing Schaffer's monist theory is that it sounds cool. The coolness just provided a kind of *catalyst* for the search for your real reasons. The **reasons for which** you believe I'll call **motivating reasons**, following philosophers of practical reason. As I'll suggest later, your motivating reasons for belief play a role in determining whether you are *doxastically justified* in believing things, and so matter greatly for normative epistemology. It is always important, though, to be careful in distinguishing between the class of explanatory reasons and the *smaller* class of motivating reasons. Neglect of this distinction often leads people to make certain claims much too hastily.[21]

It is also important to note that while motivating reasons are uncontroversially explanatory in *some* sense, it is controversial exactly *what* sense is most directly at issue. It is particularly controversial whether it is the sense at issue in the most familiar kinds of *causal explanation*. Jonathan Dancy has famously—and very plausibly—urged that we must allow that the reasons *for which* people φ can also be *good reasons* for φ-ing. Strikingly, Dancy thinks that this undermines the Davidsonian tradition of viewing motivating reasons as causes.[22] Let's briefly consider why one might think this in the *practical* case, just to mark why one might take there to be a difference between explanatory reasons in one obvious, prominent sense (i.e., causes) and the sorts of explanatory reasons that motivating reasons are. This will become important later, as I'll later be arguing for an epistemological analogue of Dancy's view.

Suppose, with Dancy, that *our* reasons for acting must be *capable* of being good reasons. What are good reasons for acting? Good reasons for acting are most plausibly just *facts about our circumstances*.[23] If someone is trying to kill your innocent mother, *that fact* is a good reason for you to interfere with them. It might be a good reason for interfering *only because* it is suitably related to the *contents* of some of your attitudes, as some subjectivist views of practical reasons would imply. But that doesn't make the *attitudes* the ultimate reasons: at most, only their *contents* are.

---

[21]Consider, for instance, people who think that chicken-sexers have reasons that justify their beliefs of which they are unaware. The reasons here are explanatory in a merely causal sense, and not motivating. This is, I believe, better viewed as a case of knowledge without justification—or perhaps merely of a *sorting ability*.

[22]Cf Dancy (2000) and (2004b).

[23]This is something that the most sophisticated subjectivists allow. As I'll note later in drawing a distinction between normative reasons and background conditions for their existence, Mark Schroeder's desire-based view does *not* claim that good reasons *just are* desires in the most fundamental case. Rather, his view is that desires *explain why* certain *distinct* things—namely, facts in the world—are good reasons.

We must allow, Dancy thinks, that facts about our circumstances can be *our reasons* for acting. Otherwise we could rarely act for good reasons. But this creates a certain interesting pressure when we think about the difference between **good cases** and **bad cases**. In a good case, the situation is as the agent assumes: someone really is, let's say, trying to attack your innocent mother. In a bad case, the situation is not as the agent assumes: you (rationally) think someone is trying to attack your innocent mother, but in fact that's not so. In the good case, your motivating reason is the fact. If asked "Why are you trying to interfere?" you say, "He is trying to kill my innocent mother." Notice that if you're (unwittingly) in a bad case, you would say *exactly the same thing*. This suggests that *your reason* is the same in the good and bad cases.

Yet how could that be? It is a fact in the good case, and not in the bad case. The natural thing to say is that your reason is really a *proposition* in both cases. Since a fact *just is* a true proposition, this does not undermine what we want to say about the good case.[24] This proposal is *equivalent in effect* to Dancy's, though he here dislikes talk of propositions (for badly overgeneralizing reasons).

This leads him to deny that motivating reason ascriptions are *factive*. As he emphasizes—plausibly, in my mind—this is not problematic on its face. The following is not incoherent:

> Jones's reason for doing it was that it would increase his pension. Sadly, he was mistaken about that.

If you tell a nice story about a bad case involving Jones, saying something like this can be perfectly acceptable. This is strikingly *unlike* paradigmatically factive contexts, such as "sees that". It would *always* be incoherent to say, "John sees that there is a bear over there, but there is not a bear over there." Yet sentences that express *causal* explanations *are* factive. The sentence "A happened because B" entails that B occurred (as well as A). This leads Dancy to think that motivating reasons—which are contents of beliefs—aren't (always) explanatory in a *causal* sense.

Instead, they are explanatory in a *normative* sense: we **make sense** of what someone is trying to do by considering the considerations that led them to regard their action favorably.[25] These *considerations* may or may not be facts. Of course, they must be in good cases. That is how we can act for good reasons. But symmetry considerations pressure saying that false considerations can be our reasons in bad cases. Hence the discrepancy between the two types of explanation. Dancy and others—e.g., Millar (2004)—thus suppose that there is just a different kind of explanation. It is the kind whereby we try to render **intelligible** someone's behavior by finding the favorable light in which that person saw their situation. Motivating reasons would be explanatory reasons in *this* sense, on their view. And explanatory reasons in this sense are not (always) causes. They may

---

[24]Of course, Dancy himself doesn't like the idea that reasons in bad cases are propositions, because he thinks it is odd to say, "My reason for A-ing is the proposition that P". This, however, is widely recognized to be an error in Dancy's thinking. Notice that it is also odd to say, "I hope the proposition that P". This has not prevented people from thinking that propositions are the objects of attitudes like hoping. What this suggests, in turn, is that there is just a much more general *linguistic* puzzle—one, indeed, that some—e.g., King (2002))—have solved, compatibly with the common view that propositions are objects of attitudes.

[25]Cf. Millar (2004), who claims that explanations that cite normative facts are central to the workaday business of "understanding people": "To understand people, we must take account of the fact that they are sometimes aware of what there is reason for them to think and do […]. And, because they have such awareness, considerations of these sorts can have a role in the explanation of what they think and do. Sometimes people believe there is reason for them to think this or do that, *and the explanation for why they so believe is that there is reason for them to think this or do that and they have managed to recognize this*." (23–4; my italics)
   I would agree with much of what Millar says in this connection, though I'd add that normative factors can play an explanatory role even when we lack beliefs about them: in the good case, when you reason in accordance with a rule, we make sense of the belief you form by appealing to the correctness of the rule *even if* your complying with the rule is not accounted for by a belief that the rule is correct. (Perhaps it is explained by an *assumption* that the rule is correct. But that would be a very different story.)

always be *indirectly* linked to causes: they are the contents of the attitudes (e.g., beliefs) that cause behavior. But they are not *themselves* causes, as Davidson had it.

I shall return to the issue of how it is that motivating reasons are explanatory reasons at various points later on. The Dancyean points generalize, I believe, to our thinking about motivating epistemic reasons—the reasons for which we believe what we believe—as I'll argue in §2.3. For now, it is just important to see that the claim that motivating reasons are explanatory reasons does not trivially commit us to viewing motivating reasons as causally explanatory reasons.

So much, then, about the motivating/explanatory link for the moment. Thinking about it has revealed *two* senses of 'reason' worth acknowledging. But the third and most important sense is the one I've left for the end. This sense of 'reason' is found in claims like "There are no good reasons for believing P" and "You have sufficient reasons to believe Q". The epistemic reasons here are **normative reasons**. They *count in favor* of certain responses (e.g., belief, disbelief, suspension), usually, in the epistemic case, in ways that turn on their relations to the truth or falsity of the contents of these responses (or of related responses, as in the case of suspension).

Normative epistemic reasons shall be my main focus as I now go on to discuss the relation of reasons to other normative facts in epistemology. But I will also return to Dancy-inspired questions about *how* motivating reasons *for belief* could also be good epistemic reasons.

2.1.   *Reasons there are, reasons had, and background conditions that provide reasons for one*

In thinking about normative epistemic reasons and how they provide a reduction base for other epistemically deontic facts, it is crucial to distinguish two questions:

(Q1)   What (normative) epistemic reasons *exist* for you to believe P?

(Q2)   What (normative) epistemic reasons do you *possess* for believing P?

These are distinct questions simply on general grounds. For, generally speaking, not all normative reasons for someone are possessed by her. In thinking about what this difference comes to in epistemology, we will see that it is also necessary to distinguish (Q2) from:

(Q3)   What reasons are *provided* (by certain factors F_i) for you to believe that P?

To bring out all the distinctions, consider a point of comparison. In the practical sphere, the difference between (Q1) and (Q2) is clear. Suppose a lion behind Door A would eat you. The fact that this lion is behind Door A is a conclusive reason for you not to open it. Crucially, it is not such a reason for *just anyone*. This lion may love others, and just tranquilly purr at them. But if you haven't the faintest clue that the lion is behind Door A, you may not *possess* this reason.[26]

---

[26]Skorupski (2011) nicely makes this point with another case: "There is a reason for Mary to slow right down as she approaches this bend, over and above the usual reasons for caution around bends. A very large truck has broken down just the other side of it and is blocking the whole road. However, Mary cannot tell that that has happened until she has got far enough around the bend to see the truck, by which time it will be too late to avoid hitting it. It is only when she reaches that point, where she sees the truck, that she *has* a reason […] to slow right down. Even so, there *is* a reason for her to slow right down before she gets to that point; namely, the fact that a truck has broken down round the bend. This becomes a reason *for her* to slow down as soon as she gets to the earlier point at which slowing down is necessary to avoid hitting the truck. We could […] put up a warning triangle at that point." (108, some italics mine)

Some resist this idea even in the practical literature. Schroeder (2008) says that it is a mistake to think of having a reason as a complex state factorizable into one's bearing a *possession* relation to that reason—a view he calls the "Factoring Account". He suggests that 'have' as it appears in these contexts functions

So, there is a decisive reason for you—and not for everyone—that you nonetheless do *not have*.

Is there a similarly clear difference in the epistemic case? One might venture an emphatic 'yes' with the following speech: "Suppose P is true and entails Q. Here *that P* is a sufficient reason for believing Q. But if you are irremediably in the dark about whether it really is a fact that P, you need not *have* any reason to believe Q. What could be clearer? Isn't *this* the difference we want?"

Not clearly. In the case in this speech, it may not merely be true that you *don't have* a reason to believe Q. It is also unclear that there *is* a reason *for you* to believe Q. That is what we want. We saw this structure in practical case: there, there was a reason *for you* not to open Door A that you didn't *possess*. I am *very* unsure that the epistemic case we've just considered is exactly analogous.

Let me explain. If P is just some extra-mental fact, then even if it does entail Q, it does not follow that there exists a reason *for one* to believe Q. Setting aside other logical or probabilistic routes to Q, there would only exist a reason for one to believe Q if there existed a reason for one to believe P. Yet there is not a reason—not even *pro tanto*—for one to believe *any* truth under *any* conditions. A reason to believe P is a real *pro tanto* license to believe P. This is an instance of the general principle that a reason to φ is a genuine *pro tanto* license to φ. There existed a real, undefeated *pro tanto* license for you to refuse to open Door A in our practical case. (Indeed, you were *prudentially required* to refuse to open Door A, though *given* your lack of access to the relevant fact, you would have been *irrational*.) Getting back to the epistemic case, the mere obtaining of an extra-mental fact does not always ensure that there is a license—not even *pro tanto*—for one to believe it. Otherwise there would be *pro tanto* licenses *for us* to believe *any truth*.

So what is going on in the case in the speech? The initial claim was that the fact that P itself yields a *detached reason* to believe Q. Not so. What is so is that there is a *wide-scope* reason,

$$R[B(P) \rightarrow B(Q)], \qquad \text{for anyone,} \qquad \text{constituted by the fact that P entails Q,}$$

as opposed to R(B(Q)).[27] This fits with what I say. Only if R(B(P)) holds for one can *deontic* detachment apply to yield R(B(Q)), which is what was said to be true in the speech. *Factual* detachment fails. From P and R[B(P) → B(Q)], for anyone, it doesn't follow that R(B(Q)) for one. Compare the deontic logician's old point. If you're going to murder me, you should at least do it gently. It doesn't follow that *since* you are going to murder me, you should murder me, and gently. One explanation of why this is false is that the first 'should' scoped widely over 'if' at the level of logical form. Something similar goes in some specifications of our epistemic case.[28]

---

pleonastically, like when we talk of *having partners*. We do not analyze your having a partner as there being a partner to whom you stand in a *having* relation. Rather, having a partner is just a matter of bearing the *partnership* relation to a person. So Schroeder suggests a dualist account on which there are really two *reason-relations* to account for the distinction: (i) the subjective reason-relation, which you bear to a state of affairs you *take* to obtain that would, if it obtained, favor acting in some way, and (ii) the objective reason-relation, which you bear to a fact that does favor acting in some way. Schroeder would claim in my case that you lack a subjective reason to refuse to open Door A but have an objective reason to refuse to open Door A.

    I think he is wrong for reasons stressed by Lord (2010), who presents a version of the Factoring Account, gives independent arguments for it, and convincingly explains why Schroeder's objections fail.

[27]See Broome (1999) for a discussion of such reasons, and of their importance in making sense of theoretical and practical reasoning. Broome also thinks that requirements of rationality are always wide-scope reason-relations. But one should distinguish the weak claim that *there are wide-scope reason relations*—which helps me in the case I'm discussing—from the *theory of rationality* according to which requirements of rationality are *all* wide-scope. I do not need to invoke anything this strong in the case I'm discussing. Indeed, I'm not even discussing rationality yet. So what I say above shouldn't be terribly controversial.

[28]One may also explain this with a conditional 'ought'. Then the idea would be that there is a non-detaching conditional reason in our case. The difference here is not important, since the upshot for us is the same.

At best, a fact is a *pro tanto* license for belief *only when* further facts about the people for whom it is to be a reason hold. The task of figuring out what these *further facts* are is a large part of what epistemology should be about. I won't take a stance yet. But there is a familiar array of options: if the fact that P is a worldly fact, these further facts could be at bottom facts about people's *non-factive experiential states*, about their *factive mental states*, or about their *available belief-independent belief-forming processes*, or so on. These further matters **provide** reasons for people to believe things.

For a factor F to provide a reason for S to believe P is for F to make it the case that some fact—often a different, typically worldly fact—is a reason for S to believe P. This further factor F *confers the status* of being a reason onto a fact—or maybe a mere proposition, on some views—which itself bears a suitable (logical or probabilistic) relation to P. As I will stress again in §2.3, this is not for F *itself* to be the real reason or even part of the real reason. For as Mark Schroeder has rightly insisted, we can distinguish between reasons and **background conditions** on the existence of something *qua* reason—conditions that *explain why it is a reason*.

Mental factors often explain why facts are epistemic reasons for people. For one thing to explain why some other thing *is what it is* is not necessarily for it to be *part* of that latter thing, or a thing of the *same relevant kind*. This idea fails miserably in other cases, as Schroeder noted.[29] So it is nontrivial to collapse background conditions on reasons into reasons. Schroeder's own desire-based account of *practical* reasons illustrates how a divorce could work (even if the theory is itself false). His view was not that reasons *are* desires, but rather that there must be some desires suitably related to a fact in order for that fact to constitute a reason. Desires are the background conditions that *explain why* other things—worldly facts—are reasons. While I'd reject desire-based views, the illustration helps. I recommend a structurally analogous view for understanding what it is for there to be epistemic reasons for people. Mental states and processes are the *background conditions* for facts to be epistemic reasons, and are not themselves these reasons.

I will discuss this framework more in §2.3. For now, simply suppose that a factor F provides a reason for S to believe something, understood along these lines. What matters now is this. It is *not yet* for F make it the case that S has this reason. That something provided a reason *for you* to φ is not yet in general for you to *possess* it. That was clear in the practical case. There, all the background conditions for R to be a reason for you to φ can hold—so that R can be a reason for you—and yet you could fail to possess R. The difference we are looking for is not between reasons *sans phrase* and reasons for people. Reasons are always reasons for people, though typically in virtue of various (often distinct) background conditions. Facts that do not bear such a relation are mere facts. So the earlier speech fails even more deeply.

When we see what the distinction between existing and possessed reasons *can't* come to in epistemology—it can't be conflated with the distinction between reasons *for you* and bare facts related by wide-scope reason-relations—it is less obvious that there is such a distinction in the

---

[29]"[The claim that there is no distinction between reasons and their background conditions] denies a distinction with respect to reasons that is a perfectly good distinction in many other domains. Although being inaugurated is a necessary condition for someone to be president of the United States, and the fact that George W. Bush was inaugurated less than four years ago is part of why he is president, the fact that Bush was inaugurated isn't part of the president of the United States. It is merely a background condition on George W. Bush being president of the United States. Similarly, in order to be a piece of corn on the cob, a vegetable must have grown on a maize plant. But the fact that this particular yellow, buttered, and salted tasty morsel was grown on a maize plant isn't part of the piece of corn on the cob that I'm eating—the piece of corn on the cob is just the yellow, buttered, and salted tasty morsel itself. So in general we have no trouble distinguishing between the facts necessary in order to explain why something is an F, and what the thing is, which is the F." (Schroeder (2007: 24))

epistemic domain. But we must on reflection accept it. Here is a case that illustrates it. Suppose S knows P, and P entails Q only by an extremely intricate inference S could never make out, and that there is no other reason for S to believe Q. Does S have reason to believe Q? If we are hearing this we ought, we'll say 'no'. Still, S's knowledge that P does provide a reason to believe Q. We should not deny that. It is just that S doesn't *catch a hold of this reason* (*qua* reason for believing Q), because he is blind to the inferential route. So S possesses no reason to believe Q.[30]

This is the case we're looking for. If this sounds impossible, it is only because we are conflating two subtly different thoughts:

> (a)   There is a reason for S to believe Q (provided by his knowledge of P),

and

> (b)   S has a reason to believe that Q (namely, P, which is had *via* his knowledge of P).

(a) is plausible and true. But it is irrelevant. For we must deny the inference from (a) to (b). There are principled reasons for doing this. The more general pattern of which this inference is an instance is clearly invalid. The more general pattern is just the pattern from:

> (c)   There is a reason for S to φ (provided by F),

to

> (d)   S has a reason to φ (in virtue of F)

It was clear from the beginning that (c) does not entail (d). This was one point from the case of you and the lion behind Door A. You shouldn't open Door A, since a lion would eat you, and there are no reasons for you to want to be eaten. So there was a conclusive reason *for you* not to open the door: namely, that this lion was behind it. Once again, this fact is not a reason *for anyone* not to open Door A. The lion may befriend others. Still, you didn't *possess* this reason-for-you.

Of course, one could try to claim that what is *ultimately providing* the reason is not just the worldly fact, but something about you—say, your desire not to be eaten, as Humeans about practical reasons might say. But even if we went Humean in this way, we would *still* need to draw a distinction between (c) and (d). We could claim that the desire is just a background condition that provides a reason, and that the reason provided is the fact that the lion is behind Door A. That, as it happens, is probably what the most sophisticated desire-based theorists say. We need these distinctions in the practical case, and we also need them in epistemology.[31]

Later in §2.9 I shall discuss further the difference between being provided with reasons and having the reasons that are provided. I will also explain what I think having reasons comes to in

---

[30]Here is another analogy to hit home the point that simply knowing some fact, which is a reason for your A-ing, is not to possess a reason for you to A—though it is to have access to a fact that is such a reason. Suppose you know pressing a button would open Door A. Again, there is a lion that would eat you behind Door A. But you don't know this. The fact that pressing this button would open Door A is a conclusive reason not to press it. You do know *this* fact. Still, you don't possess a reason not to press the button.

[31]Many epistemologists seem to miss this point in spite of seeing the distinction between reasons that exist for people to A and reasons that are had. Consider Pryor (2005)'s apposition: "For something to be a reason you have, for it to justify *you* in believing P, it has to be in some sense epistemically available to you." And consider BonJour (1985: 101): "What we must now ask is whether and how the fact that a belief coheres in this way is cognitively accessible to the believer himself, so that it can give *him* a reason for accepting the belief." The italics mislead in both cases. Reasons are *always* reasons *for people* to respond in certain ways. A reason can be a reason for you to A even if you don't have it. There is nothing in the notion of an agent-relative reason— a reason *for some person*—that entails *possession* in the relevant sense. This is evident in the practical case. We ought to say the same kind of thing in the epistemic case.

epistemology. For now, it is enough to see that these distinctions exist.

2.2. *Getting connected with good reasons*

Having sufficient reasons is one thing, believing for sufficient reasons is another. This distinction is like the familiar one between propositional and doxastic justification. While I think these distinctions are the same, it is controversial to think so. To see how they could be even close to being the same, some words are in order about what believing for a sufficient reason is.

I start with parallels in other domains. It is, I think, a general fact that there are weaker and stronger ways of understanding what it takes to *φ for a sufficiently good reason*.

Consider the following case to see the point in the practical sphere. Bill is obliged to give a gift to Mary. He can press one of several buttons to send one to her. Being perversely egocentric, he will just buy whatever he would want. So he hits the button that will send her some LP by David Bowie. If asked why he pressed that button, he would say: "C'mon, it's *Bowie*. What more reason could you want? If Mary is crazy enough not to like him, she doesn't deserve a gift anyway." As it happens, Mary has just become a Bowie fan, but she doesn't yet own any of his records. Indeed, there is no button that would send her something she would enjoy as much. Bill is even in a position to know this. But he completely ignores the fact in choosing.

Does Bill act for a sufficient reason? Here is one argument for saying 'yes'. Bill's reason for pressing the button is that it will send Mary a David Bowie record. Since Mary has just become a Bowie fan and will get more enjoyment from this record than from anything that would be sent to her if he hit other buttons, this is a sufficient reason for pressing the button. So, Bill's reason is sufficiently good, and he acts for it. It follows that he acts for a sufficiently good reason.

Still, this conclusion sits uneasily. In first hearing the question at the head of the last paragraph, I wanted to say 'no'. Rather than flat-out embracing the conclusion, I find it better to distinguish a weak from a strong understanding of what it takes to φ for a sufficiently good reason. On the

> *weak understanding*, one φ-s for a sufficiently good reason iff there is some R such that R is in fact a sufficiently good reason for φ-ing, one φ-s, and R is one's reason for φ-ing.

Since Bill's reason for pressing the button is that it will send Mary a Bowie record, and this is indeed a sufficient reason, Bill acts for a good reason on the weak understanding. But on the

> *strong understanding*, one φ-s for a sufficiently good reason iff there is some R such that R is in fact a sufficiently good reason for φ-ing, one φ-s, and one's having R as one's reason for φ-ing is *explained by R's actually being a sufficiently good reason for φ-ing*.[32]

---

[32]You might resist my presupposition in stating this claim that normative facts could play a role in explanation. I think this would be a mistake. Others have done much to argue that there is nothing confused in this idea; see Wedgwood (2006) for a lengthy, excellent defense.

But I really don't need anything as bold as Wedgwood's view. One could hold that having R as your reason is explained by R's actually being a good reason when your having it is explained by what makes that reason a good reason, *where* this proceeds by the triggering of a sensitivity that you have to good-making features of this kind (even if you do not conceptualize these as such). In general, the facts that *make* reasons good are not themselves normative facts. So what is doing the explaining here needn't *in the first instance* be the normative fact. The normative fact could then do the explaining in virtue of the subvening facts doing the explaining. Of course, some might say that this would lead to overdetermination, that this is intolerable, and that the subvening facts thus actually screen off the fact they subvene. But I'm with Schaffer (2003) in thinking that such overdetermination is utterly commonplace and unproblematic.

Of course, Bill assumed that his reason was sufficiently good. But notice that if it had *not* been a sufficiently good reason—if Mary had hated Bowie—he would have done exactly the same thing. This suggests that the actual sufficiency of this reason had nothing to do with its being his. There is a difference between the *merely supposed* sufficient goodness of a reason explaining why it is yours, and its *real* sufficient goodness explaining why it is yours. Since acting for a sufficiently good reason in the strong sense requires the latter, Bill did not act for a sufficiently good reason in this sense. That, I suggest, is why we persist in wanting to say 'no' to the question asked earlier, even though an argument can be given that Bill acts for a good reason. We are tracking a stronger notion.

I should stress that I am not making big assumptions about what it takes for the fact that a reason is good to explain why it is your reason. I do not assume that you have to (justifiedly) believe your reason is good in order for its goodness to be relevantly connected to its being your reason. Perhaps you are naturally a practically wise agent, and are disposed to act only if you have a good reason. This could be a fact about you that is generated by more or less innate, subpersonal mechanisms. While virtue is often a consciously acquired habit, you could just have a knack. This may be what we should say about Huck Finn, who was praiseworthy in his treatment of Jim in spite of his belief that he acted wrongly. Indeed, I'd say the most plausible story to tell is one that appeals to abilities to act on reasons of certain sorts only if those reasons are sufficiently good. Call these *reasons-sensitive abilities*. One's A-ing is explained by R's goodness iff R's being one's reason for A-ing manifests a relevant reasons-sensitive ability, where relevance is a function of the bearing of the reasons in the range of this ability on A-type acts. Having R as one's reason for A-ing will manifest a disposition to use certain reasons as one's reasons for A-ing only if they are good reasons for A-ing.

We can draw the same distinction in epistemology. On the one hand, there is

> believing that P for a sufficiently good reason on the weak understanding, which requires that one's reason for believing P is some R which is a sufficiently good reason for believing P,

and, on the other hand, there is

> believing P for a sufficiently good reason on the strong understanding, which requires that R's being one's reason for believing that P *owes to the fact that R is a sufficient reason for believing P*,

where we understand that R's being one's reason for believing P owes to R's goodness as a reason for believing P *only if* R's being one's reason for believing P manifests some relevant reasons-sensitive ability. The disposition at issue here is one to heed just those considerations that really bear on questions like the question of whether P.

Now, before I exploit this distinction, it is worth considering an alternative way of capturing the intuitions that led me to it. The alternative is simpler in one way, but too demanding in another. It comes out in the idea that Bill does not act for a good reason because the fact that pressing the button would send Mary a Bowie record is not *by itself* a sufficiently good reason for pressing it. The *real reason*—the deeper one that *explains why* this is a worthy consideration—is that Mary would enjoy what would be sent to her if he pressed this button. Since Bill doesn't attend to this in deliberating about what button to press, his reason for acting isn't sufficiently good.

We should reject the general idea of which this one is an instance. The general idea is

> (X)   that a normative reason is never complete unless it is conjoined with all the facts that make it good,

and, more crucially, that

> (Y)   one never really acts for this good reason unless one acts for these deeper reasons, or acknowledges them in one's deliberations about what to do.

This is implausible. We are often unaware of what ultimately makes certain enterprises or states of affairs good or worth producing. If we weren't, normative theorizing would be a cinch. The suggestion we are considering implies that in all these ordinary cases we fail to act for good reasons. This is not so. We often are not acting for whatever deeper facts make our enterprises worthwhile. If so, it follows that

> (Z)   we can act for a good reason R even when we don't in deliberation make further reasons of the deeper facts that make R good.

If (Z) is true, (X) and (Y) are false, and the specific suggestion considered earlier cannot afford an adequate replacement for the notion of acting for a good reason in the strong sense.

So, the initially tempting suggestion does not capture what was really amiss in Bill's case. There is nothing generally amiss with not acting for the deepest reasons—the ones that make all other reasons good. The considerations that one acknowledges in deliberating and then in φ-ing are not automatically better if they are deeper. Their goodness just has to do with whether they are *in fact* grounded, as a point of normative metaphysics, in the ultimate reasons.

Of course, it is not as if the deeper reasons are *irrelevant*. While they needn't figure in one's deliberations, or even be known, one must at least be *tracking* them (*via* a reasons-sensitive disposition). In the case we've been considering, it's clear that Bill is not tracking what makes his reason for pressing the button sufficient. What makes it sufficient is that Mary will actually enjoy the LP. Bill is not tracking this. So, we are back to my view, which, while more intricate, is better than the suggestion that would always put the good-making facts at the forefront of deliberation.

There is a simpler objection to the alternative suggestion in the epistemic case. It invites the Lewis Carroll regress. If it is true that P, and that P entails Q, then if you know that P there is a good reason for believing Q. Now on *my* view, one can believe Q for a good reason in this case simply in virtue of P's goodness as a reason for Q playing a role in explaining why one has P as one's reason for believing Q. Often the way this works is that one can just track the fact that P entails Q *by being a competent reasoner*. One is disposed to conform to a pattern guaranteed to be good by logic. This is true of smart ordinary people who are not familiar with axiomatic presentations of propositional logic. They just see the relevant inferential connections *by being competent* in making the right inferences when prompted. (Thus natural deduction's popularity.)

The alternative view refuses to acknowledge this fact. It requires a further belief in the inferential connection for one to count as relevantly tracking it. But we now tempt the regress. To the extent that one *isn't* a competent reasoner, simply having this further belief and the belief in P *result in* one's belief in Q won't suffice for one to have truly *reasoned well from* P to Q. This was the Tortoise's predicament. Instead of requiring that one track the logical connections by making them an *explicit part of one's deliberations*, it is better to insist that one track the connections by *being a competent reasoner*.[33] This is what is required for believing for a good reason in the sense that matters. And

---

[33]Millar (1991) developed an appealing account of this form. His view was that believing that *p* for good reasons *q, r* … for *p* (in the strong sense) involves "an exercise of competence on his part with respect to a set of patterns of inference which underwrites the fact that *q, r*, … constitute an unconditional reason for *p*". (59)

that sense is the strong sense. It is not the weak sense plus dubious further requirements that expand the set of reasons one must use to believe for good reasons.

One last comment is in order. You might be suspicious of the very attempt to call the strong understanding of believing for a good reason an understanding of believing for a good reason at all. Why isn't it just a *trivial truth* that believing for a good reason is believing for a reason that is, as a matter of fact, good? And why isn't it a stipulative evasion to use the strong understanding?

My reply is that these questions get things exactly backwards. The strong sense is the *dominant* sense. What is claimed to be a trivial truth may be no truth at all. φ-ing for a good reason in the sense we care about involves not just *conforming* to that reason's demands, but *complying* with them. There is a tight analogy here with following correct rules. If you do not know whether you are supposed to drive on the right-hand or left-hand side of the road, and drive on the right-hand side just because it feels nice, while you conform to the correct rule, you do not comply. We can describe what is going wrong as a failure to track what makes the rule correct. Intuitively, this implies in one clear sense—the clearest sense, I'd assume—that you aren't following the correct rule. But all that it is to comply with the demands of a reason in φ-ing is for one to have that as one's reason for φ-ing *because* is it is good. As noted earlier, the 'because' here does not in general indicate any explicit deliberative attention, nor does it indicate some *further* motivating reason. It need only require a certain kind of modally robust tracking of whatever it is that makes the reason good, ideally, I'd say, *via* a competence.

Just as the ordinary notion of following a correct rule involves compliance and not just conformity to its demands, so the ordinary notion of φ-ing for a good reason involves compliance with that reason's demands and not just conformity. If what it means to comply with the demands of a reason is for the rightness of those demands to explain why one has that reason as one's reason, the ordinary notion of φ-ing for a good reason is the strong sense. So, there is nothing suspicious about the strong sense of φ-ing for a good reason. We have as much reason to put the strong sense of φ-ing for a good reason first as we have for putting the strong sense of following a correct rule first.

2.3. *The ontology of epistemic reasons*

There has recently been some discussion of what epistemic reasons are. One sees three views:

Factualism:	epistemic reasons are facts;
Abstractionism:	epistemic reasons are propositions (or states of affairs);
Statism:	epistemic reasons are mental states.[34]

And he says that "competence with respect to an inference pattern is a matter of being governed by that pattern", where this governance is a matter of having certain "habits of belief management" and a "sensitivity to inferential links instantiating the pattern". (59, 62). He spells out in detail exactly what these habits are, and what the relevant kind of sensitivity involves. I don't know that I would require anything this complicated, but the core thought in his account strikes me as extremely attractive.

[34]I take this language from Turri (2009). Abstractionists agree with statists in good cases: they will usually say that *motivating* reasons are *propositions sans phrase,* and *normative* reasons are *true propositions* (=facts).

Turri strictly defines factualism as the view that all reasons are non-mental facts. But *this* view is too implausible to be worth considering. It is rightly held by *no one.* Any sensible factualist would want to allow that I have a reason to believe that I am in pain, and the fact that I am in pain is clearly a mental fact. The factualist simply denies that reasons are *necessarily* mental affairs. Remarkably, Turri uses the uncharitable part of his definition of factualism in objecting to a move he thinks the factualist will make in reply to his objection from the case of reasonable belief in a demon world. This is a big dialectical error.

Many epistemologists like statism, and regard factualism and abstractionism as implausible. (This very fact is, I think, partly connected to why some epistemologists think we cannot do epistemology reasons-first.) This is striking, since the analogues of factualism and abstractionism are extremely popular in the philosophy of practical reason, *even among* people who think that the *source* of practical normativity *lies within*—say, in desire, or in the will, on the two major internalist views. I think that the preference for statism in epistemology is misguided. More precisely, there is just a terminological dispute lurking here.

My claim rests on a simple point discussed a bit earlier. We can allow that there's a difference between the claim that

> (1)   Something *is* a reason for S to φ,

and the claim that

> (2)   Something *provides* a reason for S to φ.

These claims need not be equivalent. The debate about whether reasons are *mental states* as opposed to (possibly non-mental) *facts* or *propositions* becomes a lot less interesting once this distinction is drawn. Let me expand this suggestion.[35]

Consider the following simple view for the epistemic case:

> (3)   mental states *provide* us with epistemic reasons and enable us to *reason with them*,

whereas

> (4)   possibly non-mental facts or propositions *are* the epistemic reasons provided,

where, in general, I understand something's *providing* a reason for me to φ as

> its making it the case that some (often distinct) fact is a reason for me to φ.

Call (3&4) so understood the *Compromise View*.

Here is an example of how to think in the terms provided by the Compromise View. Completely inaccessible facts in the distant past are not by themselves reasons for me to believe anything. Only if I were connected with these facts in some way could they become reasons for me. That does not mean that my *connection* would now be the real reason. It is just the factor *in virtue of which* these facts would become reasons for me to believe things—hence, "the provider". It is, to borrow again the fine idea from Schroeder, a *background condition* on the existence of these reasons *qua* reasons for me. Of course, the facts already existed, but not *as* reasons for me. The Compromise View makes this type of claim across the board. When paired, for instance, with a Pryor-style epistemology of perception, it would say that experiences *provide* reasons for believing their contents, but that the reasons that are provided are, at least in the good cases, *facts*.

The Compromise View lets us capture the key things that statists, abstractionists and factualists all want to say, and that these parties often believe they cannot all legitimately say. It captures the

---

[35]Williamson (2000: 197) anticipates this suggestion: "Experiences provide evidence; they do not consist of propositions. So much is obvious. But to provide something is not to consist of it. The question is whether experiences provide evidence just by conferring the status of evidence on propositions."

"linguistic phenomenology" to which non-statists advert—e.g., it allows us to read as literally true claims that seem true like: "Her reason for believing that Jones was the murderer was that his fingerprints were on the knife."[36] More importantly, it also allows us to capture the non-linguistic phenomenology of believing for reasons, which I think affords a very major consideration in favor of non-statism. Given the familiar *transparency of experience*, we only very rarely think about our experiences. Usually, we just "fall through" to their contents. It would be astonishing if we only very rarely thought about the normative reasons that justify us when we think correctly. The Compromise View avoids this conclusion. Yet it also enables us to capture the statists' thought that mental states are what we *reason with*, and that they *provide reasons*.

Since my Compromise View inherits the virtues of these opposing views, and these opposing views do not have the same virtues, there would seem to be more positive reason to accept it than to accept any of these views. This puts it on *prima facie* strong footing.

Before endorsing it, though, I must stem some worries. First of all, it's worth noting that the view inherits an apparent awkwardness in how non-statists handle some properly basic beliefs. We can see this by reflecting on a switch from good to bad cases of perceptual belief. Suppose I look at a red sphere and, as a result, believe that there is one. The Compromise View does have the advantage of explicitly saying that my *looking* at the red sphere is what *provides* me with a reason for believing that there is a red sphere. Mental states, on this view, can provide reasons. But what *is* the reason provided? On the simplest way of developing the Compromise View, the reason that is provided is the *fact* that there is a red sphere before me. Now, there is nothing immediately bizarre in this suggestion. It is embraced by direct realists like Brewer and McDowell who claim that perceptual experience in the good case *opens us up* to the facts, and enables *these facts* to be our reasons for belief.[37] These people plausibly claim to be on the side of common sense. But things get puzzling when we turn to the *bad case*. Suppose the demon causes me have an experience with the same feel. What now is the reason my experience provides? If an apparition in the demon world asks me why I believe what I believe, I will still *want* to say: "Look: there's a red sphere *right there*." But, had I the mouth, what I'd say would pick out no extra-mental fact. If my attempt to cite my reasons is the same in this case and the good case, it is hard to see how *my own* motivating reason could differ. I still cite the same *consideration*.

If the Compromise View is a compromise, it cannot switch from the mere claim that my experience *provides* the reason to the stronger claim that my experience *is* the reason. To do so would be to retreat to statism.[38] So it looks like the Compromise View must say something a bit

---

[36]"Linguistic phenomenology" was J. L. Austin's nice alternative term for ordinary language philosophy, though I've seen others—John Hawthorne, I believe—use it simply to mean "how things seem, taking ordinary language at face value". I use it in this latter, even more useful sense.

    See Austin (1970: 182) for the original use: "In view of the prevalence of the slogan 'ordinary language' and of such names as 'linguistic' or 'analytic' philosophy or 'the analysis of language', one thing needs specially emphasizing to counter misunderstandings. When we examine what we should say when, what words we should use in what situations, we are looking again not *merely* at words (or 'meanings', whatever they may be) but also at the realities we use the words to talk about: we are using a sharpened awareness of words to sharpen our perception of, though not as the final arbiter of, the phenomena. For this reason I think it might be better to use, for this way of doing philosophy, some less misleading name than those given above—for instance, 'linguistic phenomenology', only that is rather a mouthful."

[37]The metaphor of "openness" comes from Brewer (2002). McDowell often talks in related ways. Consider: "[W]hen all goes well in the operation of a perceptual capacity of a sort that belongs to its possessor's rationality, a perceiver enjoys a perceptual state in which some feature of her environment is *there* for her, perceptually *present* to her rationally self-conscious awareness." (McDowell (2011: 30-1)) I learn that Ian Schnee has recently been developing and defending a related factualist view. See his (ms₁) and (ms₂).

[38]A rather extreme disjunctivist version of the Compromise View might endorse the switch without engaging in a retreat in the good case. But such disjunctivism strikes me as making a complete mystery out of what

odd in the bad case. Since the old fact is unavailable, and adverting to a state would give up the compromise, all we have left is the *proposition* that would be a fact if the subject weren't deceived.

Is this a problem? Well, I do not find it bizarre *on its face* to say that a proposition can be a motivating reason. Jonathan Dancy's discomfort with this claim is, I believe, parallel to the clearly misguided discomfort that some people have about propositions being objects of desire that comes from the infelicity of saying things like "I desire the proposition that P".[39] I am also not convinced that "S's reason for A-ing is that P" is semantically factive, so that it is only true that S's reason for A-ing was that P if P is true.[40] There are cases in which it can sound fine to say things of the form, "John's reason for A-ing is that P. Sadly, he's mistaken about that." Of course, I'd like to be *able* to say that when the reason is good, P will be true. The Compromise View shouldn't (and doesn't) make that impossible. But what we've learned about what the Compromise View must say in the bad case does not make this impossible. The reason can still be a fact in the good case, since facts *just are*, on a widely held view, *true propositions*.

So I can see only two remaining concerns about what the Compromise View leads us to say in bad cases. The first is that a false proposition cannot be a good normative reason, so the subject in the demon world will on the face of it not believe for a good reason. So, the view faces the "new evil demon problem". In reply, we will get a small preview of what is to come. The idea that φ-ing rationally and φ-ing for good reasons come apart is widely accepted in the philosophy of practical reason. The corresponding view should, I believe, be taken just as seriously in epistemology. This much will be argued at great length in Chapters 2 and 3. If one drew this distinction, one could apply it here, and say that while the subject is rational in believing what he believes, he lacks a good reason for believing what he believes. Obviously, the subject in the demon world is not open to rational *criticism*. He may even be rationally *praiseworthy*. We should not conflate this *hypological* intuition with the *deontic* intuition that he believes for a good reason, since these things come apart. As we'll see at greater length later, there must be a quite general difference between being rational (or irrational) in A-ing and having good (or bad) reasons for A-ing.[41] It would not be at all surprising to find this pattern in epistemology.[42]

If one doesn't like this idea in this context, one could reject the claim that false propositions cannot be good normative reasons. We might usefully think of the reason in the bad case as an *apparent fact*, and claim that if P is an apparent fact for S, S has a good reason to believe that P. That doesn't sound totally crazy. It is *barely different* from what many internalists *already* claim in claiming that nonveridical nondoxastic appearances can provide sufficient reasons for belief. Moreover, people

---

epistemic motivation amounts to. While we may not want to agree with internalists that the normative reasons that are available are *equally good* in the bad case and the good case, surely we do want to say that the subject's motivating reasons are the same. At most, the difference in the good and the bad case is just with respect to the *quality* of these selfsame motivating reasons.

[39]Mark Schroeder made this point in commenting on some posts on PEA Soup. Philosophers of language have explained the oddity of claims like "I desire the proposition that P" in a way that does not require something other than propositions to be objects of desires. Cf. King (2002).

[40]Matt McGrath and Mark Schroeder had good things to say about this in comments on a post by Clayton Littlejohn here: http://peasoup.typepad.com/peasoup/2010/02/thoughts-as-motivating-reasons.html

[41]It is revealing that Stewart Cohen, who gave us the "new evil demon problem", slides between rationality and justification, and claims that rationality is the core notion: "'Reasonable' and 'rational' are virtual synonyms for 'justified'. […] [C]learly the important concept, the one epistemologists have been concerned with, is […] 'reasonability' or 'rationality'." (Cohen (1984: 283-4)) These claims need *a lot* more defense.

[42]It is worth adding that the category of excusable wrongdoing is extremely familiar in ethics, and that some people go even farther and assume that one can even be praiseworthy in acting impermissibly or blameworthy in acting permissibly if one's evidence is sufficiently misleading or one's motives are sufficiently dissociated from the deontic objective character of one's act. See for instance Zimmerman (1997). The general idea can also be invoked here. Cf. Littlejohn (2009), Engel (1992), and Dancy (1992).

who refuse to distinguish between φ-ing rationally and φ-ing for a good reason will be pressured to accept this kind of claim in the practical case as well. We should say in Williams's old case that Bernie's reason for taking a sip is the apparent fact that there is a gin and tonic in the glass. But this is a *merely* apparent fact. If you think Bernie acts for a good reason here, you are compelled to think that a false motivating reason can be a good normative reason.

So the ultimate disagreement has nothing to do with the truth of the Compromise View, but instead with the fact that everyone will face a dilemma between either allowing that states with false contents (or the false contents themselves) can provide (or be) reasons or allowing that φ-ing rationally and φ-ing for a good reason can come apart. This is a completely general problem which people will face in reflecting about the practical case.

Let's turn to a second concern that kept me in the statist camp for a long time. It is, I think, ultimately confused. Here is one version of the worry: it may look *circular* to model foundational cases as *S believes that P for reason R*, where R is a proposition. For P = R in these cases. Isn't it problematic to believe that P for the reason that P?

Circularity cannot be the core worry. How else could we model cases of knowledge of one's own mental states? There is nothing problematic in believing that one is in pain for the reason that one truly is in pain. When you have that fact right there before you, why not appeal to it as your reason for belief? It is not as if you have no reason for believing what you believe. You have a reason *of the clearest kind*. Sure, maybe it's your *acquaintance* with the fact *in virtue of which* it can serve as a good reason. But that just means the acquaintance *provides* the reason, not that it *is* the reason. Once we concede that you can appeal to a certain fact as your reason for believing that it obtains *only because* you are acquainted with it here, the superficial circularity here must ultimately be unproblematic. We can view cases of basic knowledge about non-mental affairs in a similar way. McDowell, Brewer and other direct realists presumably do precisely this. Of course, that is a substantive dispute. The point is that there is no *inherent* flaw in the *structural* view about how foundational knowledge works suggested by my Compromise View.

So what *is* the worry? It might at bottom be that it just sounds wrong to talk about reasons in foundational cases. When I know that there is a cup before me by seeing it, I just *take it in*: that belief just comes to me through perception, and thereby amounts to knowledge. It is knowledge *through* seeing, not knowledge *reasoned* into being *from some seen fact*. But this is no objection. Something can be your reason for φ-ing without your having *explicitly made* it so, or having *explicitly reasoned* from it. Moreover, the brief speech we just considered was made from the first-person perspective. Things look different when we go third-personal and think about understanding others' beliefs. Suppose you at first know nothing about Jones. You are told only that he thinks there is a pink jaguar attacking him. A wild thing to believe! He must be crazy. But you then are allowed to see him. A pink jaguar *is* attacking him. *There's* the reason, you'd say.

So, the worry is insufficiently supported by bare intuition. The opposing intuition hits me when I look at the third-person case, though I admit that I feel the first-person phenomenology. But even if we took the first-person description of foundational knowledge at face-value, it would only reveal something about our own occurrent deliberation. We already knew that this underdetermines what we ought to say about our motivating reasons, even in epistemology.

So, I see no worry yet. Since the Compromise View faces no serious objections and captures the motivations of all the other views that fail to appreciate the difference between something's *providing* a reason and something's *being* a reason, we have sufficient reasons to accept it.

One more comment is in order about how the Compromise View could generally work in

foundational cases. In every case where foundational knowledge can plausibly be claimed to exist, there will be a *pathway* to knowledge: perception, introspection, intuition or pure understanding, certain cases of preservative memory, and so on. When these pathways lead where they ought, they *put us in touch with the facts*: facts about our present environment, about ourselves, about logic, or about the past. Our contact with these facts makes them into reasons for us to believe things, including—sometimes—the proposition that these facts obtain. The pathways *provide* reasons for belief. But the reasons are just the facts. We can thus generalize what we said about the perceptual case.

Of course, there are bad cases where these pathways mislead. In these cases, our motivating reasons—the reasons *for which* we believe—are still the propositions which would have been these facts if they were true. It is not as if our motivating reasons for belief *shift* without our being able to tell. On one view that I shall take seriously, we just lack good normative reasons in bad cases because these propositions are false. Since it remains true that if these propositions had true, we would have had sufficient reasons to believe, we still believe rationally.

This is like the way in which many think about the gap between φ-ing for good normative reasons and φ-ing rationally in the practical sphere. High-level unity between normative domains is something we should *ceteris paribus* expect. There must be something that makes all these domains fall under the same general headings. It may in part be this kind of unity. But if we must, we could also cling to the other view, and say that our reasons in these cases are *apparent facts*. If we do, we will, I believe, be obliged to do it generally, and to reject the distinction in the practical sphere too, and claim that Bernie does act for a good reason in Williams's case.

2.4.   *Epistemic reasons and oughts, permissions and requirements*

One advantage of taking reasons to be the fundamental things within the deontic sphere of any normative domain is that they come with different *weights*. By weighing up the epistemic reasons, we can illuminate the structure of various other deontic facts, such as facts about what doxastic attitude one *ought* to have, or is *permitted* to have, or is *required* to have in various cases.

How does it work? Note first that, in general, it is the comparison in strength between the reasons for a given option and the reasons for the *alternatives* to that option that determines whether one has *sufficient* or *insufficient* reasons, *decisive* or *indecisive* reasons, and so on, for the option. It will be useful to spend a moment reflecting on what the sufficiency and decisiveness of reasons amount to in epistemology, with this observation about alternatives in mind.

Besides disbelieving P, suspending judgment P is an alternative to believing that P. So, it may seem that a complete analysis of sufficiency of epistemic reasons would have to look like:

> (SRB)   S has sufficient epistemic reasons to believe that P iff the epistemic reasons S has for believing P are *weightier* than S's epistemic reasons for disbelieving P and for suspending judgment on P.

Yet one might think this is prolix. Aren't reasons for suspension *derivable* from the relations between reasons for belief and reasons for disbelief? Can't we just avoid mentioning them, and talk just about the relations between reasons for belief and reasons for disbelief?

One naïve proposal about the derivativeness of reasons for suspension would be this:

> (SRS)   S has sufficient epistemic reasons to suspend judgment on P iff the
>   epistemic reasons S has for believing P and disbelieving P are *equally weighty*.

If the naïve proposal were true, we could simplify the picture of sufficient reasons for belief:

> (SRB*)   S has sufficient epistemic reasons for believing P iff S's epistemic reasons for believing P are weightier than S's epistemic reasons for disbelieving P.

But SRS is problematic three reasons. Ultimately, I think we should simply accept a further slight revision of SRB* without relying on an argument from anything like SRS.

First, suppose one's total evidence suggests that it is indeterminate whether P. Indeterminacy is not a reason for *disbelief*. It is a reason for suspension. But SRS implies that it's false that one has sufficient reasons for suspending judgment in this case, since one's reasons for believing and disbelieving P may be equally weighty—simply by having no weight, perhaps. So SRS is false.

Secondly, it is should be an open question whether epistemic reasons for belief and disbelief are always precisely comparable. If there were cases of deep incomparability, suspension would be required. SRS will make the wrong prediction. Since it is *not* the case that one's reasons for belief and disbelief are *equal* in incomparable cases, SRS will imply that one lacks sufficient reasons for suspension. This is one of the interesting structural disanalogies with the practical sphere. In practical cases, the incomparability of reasons given by the objects of choice plausibly allows one to choose in several ways. *How* choices under witting incomparability of object-given reasons could be rational is hotly disputed,[43] but the most plausible view allows for a kind of freedom here that needn't involve irrationality. This is not so for epistemology, at least if we're really dealing with *belief* as opposed to some attitude that shares only part of belief's functional profile—e.g., *acceptance* in Bratman (1992), Cohen (1992) or Wright (2004)'s sense.

Finally, pragmatic encroachment, at least on one understanding, creates some *prima facie* trouble for SRS. One way of thinking about the stakes-related differences between cases that motivate pragmatic encroachment is that the threshold for what it takes for an epistemic reason for believing to be *sufficiently* stronger than a reason for disbelieving is determined by the stakes attached to forming the belief. If the stakes are high, the fact that the reasons for believing P outweigh the reasons for disbelieving P isn't enough: the outweighing has to be *significant enough*.[44]

To handle the first and second worries for the argument from SRS to SRB*, we can revise SRS:

> (SRS*)   S has sufficient epistemic reasons to suspend judgment on P iff S's epistemic reasons for believing P and for disbelieving P are either of the same strength, or are incomparable.[45]

But to handle the third, we also need to qualify SRB*:

> (SRB**)   S has sufficient epistemic reasons for believing that P iff S's epistemic reasons for believing that P outweigh S's epistemic reasons for disbelieving that P, and by a significant margin, where significance is (perhaps) a pragmatic affair.

---

[43]Chang (1997) is a fine anthology containing defenses and discussions of many of the prominent options.

[44]See Schroeder (forthcoming), $\mathbf{Ms}_1$ and $\mathbf{Ms}_2$ for this kind of story. The idea ultimately goes back to Owens (2000: 26), who challenged evidentialists to explain what level of evidence is needed to justify belief without appealing to pragmatic factors. (See also Fantl and McGrath (2002).)

[45]In assuming that this takes care of the indeterminacy case, I assume indeterminacy entails incomparability. I don't assume the converse, and so don't enter into a big controversy about incomparability.

This allows us to avoid explicitly mentioning suspension in our analysis of sufficiency.

We cannot avoid mentioning suspension in our analysis of decisiveness. Decisive reasons leave no room for any other option. In many cases, it is epistemically permissible to suspend judgment. If you lack interest in P, then it's fine to suspend judgment on it, even if you do *have* greater reason to believe P rather than ~P (or *vice versa*). This is why sufficient reasons for belief needn't be decisive. To be decisive, they must rule out suspension. So, we need:

> (DRB) S has epistemic decisive reasons for believing that P iff S's epistemic reasons for believing that P outweigh S's epistemic reasons for disbelieving P and for suspending judgment on P.

Whether we can ever have decisive epistemic reasons for belief is unclear. Perhaps all cases where suspension is excluded as an option are cases where that's for *merely* pragmatic reasons, so that it is *not* one's epistemic reasons that are excluding suspension as an option. The kind of weight at issue in DRB is supposed to be purely epistemic. Of course, pragmatic encroachment on weight might be true. But it might not be. If not, we may never have decisive epistemic reasons for doxastic attitudes other than suspension, only decisive non-epistemic reasons.[46]

I won't digress further on this, since it will make no difference. Even if there aren't ever decisive epistemic reasons for belief as a matter of fact, DRB would be the right analysis in principle. Given SRB** and DRB as our background pictures of sufficiency and decisiveness, we can then easily characterize oughts, permissions, and requirements as follows:

> (PERM) S is epistemically permitted to believe P iff S has sufficient epistemic reasons for believing P.

> (OUGHT) S ought epistemically to believe P iff S has strongly sufficient epistemic reasons for believing P.

> (REQ) S is epistemically required to believe P iff S has decisive epistemic reasons for believing P.

I would regard PERM, OUGHT and REQ as clear truths. This is one payoff of putting reasons first: the analysis of other deontic notions is rendered extremely straightforward.

2.5.   *Epistemic permissions without possessed epistemic reasons?*

Now, there is a worry I would expect some epistemologists to have about PERM. Consider cases of forgotten evidence. Often you can know P even if you've lost track of your original reasons for believing P. Surely, one might reasonably say, it's true that if you know P, you are epistemically permitted to believe P. So, the worry goes, it looks like you can be epistemically permitted to believe P even if you lack a sufficient reason for believing P. So PERM is false.

This is too quick. In the *clear* cases where you know, you know that P *because you recall the fact that P*. Of course, your recollection is not supervenient on any significant experience of the facts that gave rise to your knowledge. But there is still a mode *whereby* you know P. There is no reason why we cannot apply the Compromise View here, and say that your reason is given by the factive state

---

[46]I stress that this is compatible with a very different claim made with similar words. Epistemologists often call reasons for believing something "conclusive" or "decisive" when they rule out all alternative *hypotheses*. Cf. Dretske (1971) for a famous early example of this usage. If "conclusive" is used in *this* sense, I do say there are often conclusive reasons. I am simply interested in the more general kind of conclusiveness that attaches to the alternatives to the *attitude* rather than the alternatives to the *content* of the attitude.

of recalling P. The reason that's given may be the fact believed, or it may be the fact *that you are in the state*. Which way one goes depends on what we *independently* want to say, in first-order normative theorizing, about these cases.[47] My Compromise View leaves it open. Of course, it would be dialectically odd to stamp your foot if queried and say: "Damn it, I just recall that P. *That* gives me my reason." But the dialectical oddity of citing this reason needn't show that this reason is no good. Good reasons don't always make for good rhetoric.

This type of point generalizes for all cases where it can look like we lack reasons but are permitted to believe. Other cases include: basic perceptual knowledge with experience, the (supposed) basic knowledge of blindsighters, and knowledge by sheer understanding. Here one can take the view that there are more basic *factive mental states* that give reasons. The reasons given may be the facts believed, or, in some cases, they may be the facts *of your being in these factive mental states*. Which we should say depends on the particular mode we consider, and that's an issue on which the Compromise View is neutral.

Of course, earlier, I did insist that in basic *perceptual* cases, the reason is always the content of perception, whether the case is good or bad. But this, recall, was for phenomenological reasons. In switching imperceptibly from a good case to a bad case, one continues to direct one's attention at the contents of perception in looking for one's reasons for believing. One does not reason any differently, in the broad sense of "reason". Such is the transparency of experience. These phenomenological reasons may not generalize for other factive states, such as intellectually seeing that P. Maybe they do, maybe they don't. (It probably also depends on *who* we're talking about. Some people may more *literally* "intellectually see" than others.) Some, I'm sure, would find it wrong to say that your reason for believing some mathematical truth *just is that truth*, which you have as your reason *because you intuitively see it to be so*, with the seeing as a background condition. So, in this case, we might say instead that it's the fact that you are in the factive mental state that is your reason for believing. What we say is a substantive, first-order issue.

Either way, the Compromise View is fine, and so is PERM. When you have basic perceptual knowledge, you know P by seeing the fact that P. While it would be odd for enlightened blindsighters to say that they see the fact that P, there is still a more basic factive perceptual state in virtue of which they blind-see. That's what gives them their reason. When you have basic intuitive knowledge that P, you know P by grasping the fact that P, at least if we try to extend the proposal in this way. Etc. There are reasons—facts—provided by mental states. They ground the permissions to believe by being sufficient reasons. They are sufficient reasons *because* they are provided by factive states like the ones mentioned here. PERM stands untouched.

A different view is available. Whether one accepts it depends on whether one likes Williamson's idea that knowledge is the most general factive mental state. Consider how Williamson would deal with the justification of basic perceptual beliefs in good cases. Since, on his view, knowledge is the most general factive mental state, seeing that P is a way—in the sense of a *determinate* of a *determinable*—of knowing that P. So, by Williamson's E=K, *that P* will be in one's evidence set if one visually perceives that P. We have a case of "duplicate reasons" here.[48] What *provides* the reason on this view *just is* one's knowledge that P, albeit *via* the determinate *seeing that P*. Knowing provides a reason for the belief that it involves by putting that belief into a certain kind of contact

---

[47]Again, this is not a retreat to statism. *Pace* Turri's astonishingly unfair "objection" to the factualist/abstractionist, *no one* should hold the view that reasons *never* directly involve mental states. How else can one deal with introspective knowledge, such as knowledge that one is in pain? People who think reasons are facts or propositions just say that the facts/propositions *concern* one's being in mental states in these cases. If they say this in these cases, they can sometimes say so elsewhere. There is nothing inconsistent or arbitrary here. Indeed, this looks to me like the closest view you'll find to common sense.

[48]I borrow this term from Ian Schnee, who gives an insightful discussion of these cases in his Ms$_1$ and Ms$_2$.

with the reason, which is also the belief's content.

As Williamson correctly notes in replying to Brueckner (2005), the only complaint that immediately comes to mind cannot be a good complaint:

> His knowledge that his cup is red justifies his belief that his cup is red. If we want to use the somewhat obscure expression 'in virtue of' […] we can say: his belief that his cup is red is justified in virtue of his knowledge that his cup is red. Perhaps Brueckner's thought is that since knowledge entails belief, his knowledge that his cup is red *is* his belief that his cup is red (which constitutes knowledge in this case). Consequently […] his belief that his cup is red is justified in virtue of his belief that his cup is red. Thus […] his belief that his cup is red justifies his belief that his cup is red. But surely such a belief is not self-justifying.
>
> That argument is spurious. Presumably, the number 9 is composite in virtue of its divisibility by 3. If Brueckner's knowledge that his cup is red is his belief that his cup is red, then in a corresponding sense 9's divisibility by 3 is its compositeness. An inference like the one above would yield the implausible-sounding conclusion that 9 is composite in virtue of its compositeness. Such reasoning hardly refutes the harmless claim that 9 is composite in virtue of its divisibility by 3 […].
>
> If I claim that Brueckner's belief that his cup is red is justified in virtue of his knowledge that his cup is red, and that his knowledge justifies his belief, I am claiming just this: the belief that his cup is red is justified for him *because* he knows that his cup is red. The 'because' here is explanatory […]. I grant that, necessarily, if he knows his cup is red then he believes that his cup is red. Given the inference schema 'P because Q; necessarily, if Q then R; therefore P because R', we could conclude that the belief that the cup is red is justified for him because he believes that his cup is red, which is surely false. But that inference schema is clearly unsound. Suppose that I have a right to vote in the election because I am a citizen at least eighteen years old. Necessarily, if I am a citizen at least eighteen years old, then I am at least five years old. It does not follow that I have a right to vote in the election because I am at least five years old.[49]

Given the availability of Williamson's view, we can give an exhaustive response to the objection.

In good memory cases with forgotten (independent) evidence, good *a priori* cases, perceptual cases, etc., there will be factive mental states to which we can appeal that provide the reasons: recalling P, grasping P's truth, seeing P, etc. The fact *that* one is in these states may be the reason in some cases, as in the memory case and the intuitive case; in other cases—perceptual cases—the fact implied by the state may itself be the reason. If knowledge is *not* the most general factive mental state, we have straightforward knowledge-independent providers of reasons, though the reasons themselves are just the facts to be known, or the fact that one is in some factive state. This would give the defender of PERM what she needs. If, on the other hand, Williamson is right, we can adopt his view, say that it is one's knowledge in these cases that gives the reasons by putting one's doxastic state into contact with the relevant facts in some way—precisely which way depending on the *determinate of knowing* at issue—and argue that there is nothing structurally problematic by analogy. This would also give the defender of PERM what she needs. Either way, she has a way out.

So, the objection fails. Now, this approach doesn't trivially extend to *bad* cases—cases where one merely seems to recall that P, or where one merely seems to grasp that P, or merely seems to see that P. But here the dialectic from the basic perceptual case that we rehearsed in §2.3 simply repeats. A perfectly sensible view I shall consider at much greater length later is that one *isn't (as) justified* in believing in these cases, though one's belief is *(as) rational*. Alternatively, one could adopt the strategy of claiming that good normative reasons can be mere propositions—*apparent facts*. This

---

[49]Williamson (2005: 468-9). See also Williamson's reply to Brueckner (2009) in Greenough and Pritchard (2009). It is far less charitable but helps to drive home the character of the view at issue.

route is not easy to take. But there is *nothing special for epistemology* here: the same dialectic appears when we think about cases like Bernie and the proposition that there is gin and tonic in the glass. A distinction between rationality and responsiveness to objective reasons was available there. One can take the other strategy too. It just requires much heftier motivation.

2.6.   *Epistemic justification and reasons*

In general, justification is a deontic feature closely linked with permissibility. This is a conceptual truth. Only the perversity of historical dogma could lead one to find it *clearer* that epistemic "justification" is to be stipulatively defined by its functional role in the equation K=JTB+X *than* that it belongs in the deontic sphere, with "permitted".[50] How anyone could have ever been convinced that this *theory* expresses a clearer conceptual role for epistemic justification than the clear general link with permissibility is baffling. That post-Gettier tradition is independently unfortunate. It may be best—as Goldman had it in the late '60s—to analyze knowledge *without appealing to justification at all*. Indeed, this may be the best way to frame virtue-theoretic views that view knowledge as apt belief.[51] To define "justification" as having a role in a JTB+X analysis of knowledge is to rule out such theories in advance by pure stipulation. This is a mistake.

It is better to understand epistemic justification on its own terms. It is better to see it as belonging—as it does in *all other domains*—to the family of deontic concepts that include *reason*, *permission*, and *ought*. Otherwise, by my lights, one is not clearly discussing justification at all. One may be discussing "the normative status constitutively required for knowledge". But it is not a point of stipulation that this *just is* epistemic justification. Indeed, I would reject this view.

Now, on the view defended earlier, it is epistemically permissible to believe that P when one has sufficient epistemic reasons for believing P. And one has sufficient epistemic reasons for believing P, on that view, when these reasons are weightier than one's epistemic reasons for disbelieving P (by a significant enough margin). There is one type of epistemic justification that amounts to nothing more than such permissibility. It is *propositional justification*. On my view:

> PEJ=SER: S has propositional justification for believing that P iff S has sufficient epistemic reasons for believing that P.

Of course, many epistemologists do not like talk of reasons but do like talk of epistemic justification. They would take this view to be highly controversial.

Part of the reason for this is, I think, no good. It's the concern rebutted in the last section about the apparent possibility of permissible belief without the possession of sufficient epistemic reasons for that belief. If there were serious reasons for rejecting PEJ=SER, they would have to take either this form, or show that one could have sufficient reasons for believing that P without

---

[50]It certainly isn't the falsity of doxastic voluntarism that should lead one to think otherwise, as I claim in Appendix A, and suggested in the first footnote of this chapter. Even if that *were* an issue—and it would have be a *completely general* one about the applicability of deontic notions to *other attitudes* such as desire, emotive attitudes, and intention—its *being* an issue would cast doubt on whether we should be talking about *justification at all* as opposed to a purely evaluative notion, such as "being a good belief to have from the epistemic point of view". It would be a mistake to start suggesting that there is some purely evaluative notion of justification, as some—e.g., Alston and Plantinga—have done, just as it would be a clear mistake to think that there is a purely evaluative *ought-to-do* that can occur in sentences like "S ought to A".

[51]Indeed, that is roughly my view. I would then put knowledge first in the analysis of what it takes to have a reason and for something to provide a reason, along the lines of the Williamsonian model discussed in the last section. We can put knowledge first in the analysis of many other important epistemological ideas without thinking that knowledge is completely unanalyzable, as Williamson himself does.

having propositional justification for believing that P. This later case I cannot see to be possible. So I cannot see good reasons for rejecting PEJ=SER. (In a coming subsection, I shall point to another reason why many people have may avoid embracing PEJ=SER, and refuse to make reasons central normative items in epistemology. But this reason is even worse.)

Embracing PEJ=SER is one thing. It's another thing to try to reduce *doxastic* justification by appealing solely to reasons and nonnormative concepts. In §2.2, I introduced what I called the *strong* sense of believing for a good reason, and noted that it is one instance of a more general pattern whereby believing for good reasons can be understood in a stronger and a weaker sense. If we want to to reduce *doxastic justification* by appealing solely to reasons and nonnormative concepts, it is this stronger sense that we will need. I'd recommend the following view:

> DEJ=FSER: S is doxastically justified in believing that P iff S has sufficient epistemic reasons for believing that P, and believes for these *in the strong sense*.

Why is the strong sense needed? Well, apart from the fact the strong sense of φ-ing for a good reason is, as I claimed in §2.2, the *dominant* sense, and epistemology is no exception, we need it to avoid counterexamples that would pop up if we used the weak sense.

The most obvious would-be counterexamples come from Turri (2011). Now, strictly speaking, Turri's aim was to attack the traditional view that doxastic justification is to be analyzed in terms of propositional justification and nonnormative properties/relations like the basing relation, and to reverse the order of explanation by putting doxastic justification first. But he phrases the traditional view in a way that couples it with the idea that one has propositional justification for believing P iff one has sufficient reasons for believing P. So, assuming that "A's belief that P is based on reason R" entails "A believes that P for reason R", Turri is naturally read read as opposing the kind of view I like. I will assume as much, I think he'd encourage this.[52]

Here is his main counterexample.[53] Miss Proper and Miss Improper are two jurors at the trial of Mr. Mansour. They both know that (P1) Mansour intended to kill the victim, that (P2) Mansour threatened to kill the victim, that (P3) multiple eyewitnesses saw Mansour at the crime scene, and that (P4) Mansour's fingerprints were all over the murder weapon. Clearly, both Miss Proper and Miss Improper have good reasons for thinking that Mansour is guilty. And it so happens that both of them come to believe that Mansour is guilty. Miss Proper comes to believe this by noting that (P1-P4) make it objectively likely that Mansour is guilty. Miss Improper comes to believe this by noting that the tea leaves say that (P1-P4) make it extremely likely that Mansour is guilty. Intuitively, Miss Proper is justified in her belief, whereas Miss Improper is not.

This would be a problem for our approach if we had appealed to the weak sense of believing for a good reason. But it is clearly no objection to our actual proposal, which is DEJ=FSER. Indeed, rather importantly, Turri's case gives us no reason to abandon the traditional view that propositional justification takes priority over doxastic justification. For Miss Improper does not believe *for good reasons* in the dominant sense. The explanation of why (P1-P4) are her reasons for believing that Mansour is guilty is not what actually makes these reasons good. If the tea leaves

---

[52]In Turri (2011), he presents himself as making a contribution to the literature on the epistemic basing relation, but the paper is called "Believing for a Reason". This suggests he shares the assumption.

It might seem odd to talk as if this assumption is coherently disputable, but some writers use 'basing' in a broader sense that doesn't necessarily include believing for reasons. Millar (1991), e.g., seems to talk in this way. He appears to *stipulate* that basing is a causal notion, and develops his theory of what it takes to believe for a good reason by considering what must be *added* to basing in this sense. More often, though, people take the causal view to be a *substantive theory* of the basing relation, and introduce the basing relation by alluding to cases where people have certain reasons as their reasons for believing certain things.

[53]This is close paraphrase of Turri (2010: 315–6).

had suggested otherwise, she would have believed otherwise, and clearly not have conformed to what (P1-P4) recommend believing. *Complying* with reasons means conforming to them *because* they are actually good reasons. Her conformity failed to track their actual goodness. This is suggested by the counterfactual observation about what she would have believed if the tea leaves had suggested otherwise. This is how the proponent of EJ=SER* ought to explain why Miss Improper lacks a justified belief. It is what we'd expect, given the points in §2.2.

If this seems sneaky, consider an analogous case and a parody argument about the metaphysics of justified rule-following that draws on it. John is going to go for a drive. He assumes that the right way to drive is to drive however he likes. He has never consulted anyone about what the correct rules for driving are precisely because he thinks they correspond to his own preferences. He asks himself: "How should I drive?" He says: "Well, I would like to drive on the right-hand side of the road. So, that must be the right rule." So he drives on the right-hand side.

When John is exiting the driveway and going onto the street, is he justified in going on the right-hand side? We might say *he* isn't, though he is in a different sense doing the *right thing*, given the rules. If our reasoning mirrors Turri's, we might conclude from this that driving justifiedly cannot be understood just in terms of correct rules and one's nonnormative relations to them. But we would make a mistake, since the more natural conclusion to draw is that John does not actually *follow* the correct rule in the *relevant* sense. His conformity is not explained by what makes the rule correct. If he had liked to drive in another way, he would have done it, regardless of what the correct rule was. The correctness of his conformity is thus accidental.

Just as we shouldn't buy this flawed argument that driving justifiedly is prior to correct rules for driving and nonnormative relations to them, so we shouldn't conclude from Turri's observations that we must reject the traditional idea that propositional justification comes before doxastic justification, and that no additional *normative* concepts are needed to analyze doxastic justification.

2.7. *First and second-order theories and why there is no reliabilism/evidentialism dispute*

Having analyzed justification in terms of reasons and nonnormative properties and relations and parried objections, it is worth pausing for a moment to remark on what I've been trying to do.
As I noted at the outset, many people in the practical literature are interested in finding out what the most fundamental normative feature is. Reasons Basicness claims that what makes anything normative is its relation to *reasons*. This is a **second-order**, **metanormative** claim. This project is *completely different* from an attempt to specify the nonnormative conditions that *make for* the presence of normative facts. That project is the project of giving a **first-order**, **normative** theory. Part of this project is fulfilled by ethicists who argue about whether utilitarianism is true, or Kantian ethics, or contractualism, or whatever. Assuming Reasons Basicness, these people are interested in seeing *when,* in nonnormative terms, *there are* reasons for acts and intentions.

The point of the last few subsections was second-order and metanormative. I've been trying to show how Reasons Basicness could hold for one sub-domain of the domain of epistemically normative facts—the deontic ones. Seeing this is crucial for avoiding misguided worries. I have not spent much time talking about theories like reliabilism, experiential foundationalism, and so on. That is because these are *first-order* theories. These theories can, I believe, be understood to be theories about *when*, in *nonnormative terms*, we have epistemic reasons. The classic process reliabilist can be understood as claiming that we have good noninferential epistemic reasons for believing that P *only when* the belief that P would be outputted by an available belief-independent, unconditionally reliable process. The Pryor-style experiential foundationalist can be understood as claiming that we have good noninferential epistemic reasons for a perceptual belief that P *when* we have a perceptual experience with the basic phenomenal content that P. Etc. These theories

are all perfectly compatible with the metanormative claims that have been made so far.

One might find this point and the distinction behind it obvious. But epistemologists have often conflated these two projects. Consider the supposed "debate" between evidentialists and reliabilists. As I see it, there is no debate to be had here, any more than there is a debate to be had between, say, a proponent of a buck-passing account that says having intrinsic value is having intrinsic properties that give *reasons* for valuing something for its own sake, and a proponent of the hedonist view that things are only intrinsically valuable if they are pleasurable experiences. Such a debate would be completely bizarre. These theories have structurally different aims that cannot intelligibly conflict. Notably, though, the property *being evidence* is plausibly a normative property. On one view, a piece of evidence for believing P and an epistemic reason for believing that P are *exactly the same*, notionally and metaphysically. To claim that epistemic justification supervenes on evidence would then be to make a *metanormative* claim. If one thought pieces of evidence *just are* reasons for belief, it follows from (P/D)EJ=(F)SER.

That claim is obviously consistent with both process reliabilism and indicator reliabilism in *exactly the same way* that the buck-passing account of intrinsic value is compatible with hedonism. Indicator reliabilism can be understood as a view about what nonnormative conditions *make for good evidence*. Process reliabilism can be understood either in the same way, or as a view about what nonnormative conditions have to obtain for one's belief to *fit the evidence.*

Strangely, it has taken more than twenty years for this point to be widely noticed. And it is not even now recognized for what it is. Goldman (2011) and Comesaña (2010) both propose views that can reasonably be called "reliabilist versions of evidentialism".[54] Both of them talk as if they are solving some *problem* for evidentialism. The problem, as they see it, is that Conee and Feldman tell us nothing about what it takes to *fit the evidence*. If evidentialism is understood as it should be— as a *metanormative* claim—this cannot intelligibly be a criticism. One cannot intelligibly criticize someone who defends a buck-passing account of intrinsic value for failing to tell us what things, specified nonnormatively, are intrinsically good. To do so would completely misunderstand the view, which is *not* a view about the relationship between normative and nonnormative facts. Evidentialism is best viewed as a metanormative thesis of this stripe, simply on pain of a very bad view about what evidence is.

Of course, the facts that comprise the evidence are not themselves normative. But neither need be the facts that comprise the practical reasons! The fact that arsenic is poisonous is a conclusive reason not to drink it. This fact is not normative: something's being poisonous is a natural affair. Only the fact *that it is a reason* is normative. To be sure, Conee and Feldman may not see it this way.[55] But if there is any objection to them, it is that they misunderstand the character and constitutively limited ambitions of their own view.

2.8.   *Why an apparently deeper objection to the appeal to epistemic reasons fails*

I have proposed to put epistemic reasons first when it comes to understanding deontic facts in epistemology. I think the analyses of other epistemically deontic facts I've suggested are *extensionally* correct. One might, however, have a lingering worry. One might think even though it is always possible to represent the epistemological facts by talking about reasons, this talk is superfluous and unhelpful. This worry is like one about empirically equivalent scientific theories that disagree on fundamental ontology. One theory posits gravitational forces, the other theory does the work by appealing to the intrinsic structure of space-time. Could reasons be like

---

[54]One central precursor is Alston (1988).
[55]Their reply to Goldman in Conee and Feldman (2011) makes this explicit.

gravitational forces after Einstein? As I'll now suggest, the answer is 'no', as long as we are clear on the distinction between first and second-order theories of epistemic justification.

I can see two versions of the objection. The first is best brought out by returning to some earlier cases. Consider again the case of recollection *sans* mental imagery and independent evidence. It is attractive to think of recollection here as a kind of *ability*—an ability to keep track of past facts, realized by subpersonal mechanisms. How is it that one knows? Plausibly, by believing truly in virtue of this ability. The aptness of the belief just amounts to knowledge. We might fully explain knowing by appeal to abilities and the nonnormative relationships between them and the facts known. Of course, we then could try to call some of the things that figure in this explanation "reasons". But isn't this at best awkward, and in fact superfluous? Even if we could bring in reasons, doing so is unnecessary and adds nothing to the explanation. Even if a reasons-involving view were extensionally adequate, it wouldn't capture what is epistemologically important, and it would tack on unimportant details.

This objection overgeneralizes. Notice that the plausible claims here are claims about *knowledge,* and about what *explains how it is* that one has a certain kind of knowledge. Notably, we *also* have said nothing about *justification* in providing the ability-based story about memory knowledge. We could duplicate the second half of the last paragraph and systematically replace 'reason[s]' with 'justification'. We *could* claim that success *via* ability provides justification. But it wouldn't enhance the plausibility of the explanation about how *knowledge* is acquired. Indeed, I suspect this is right: people too readily take for granted the assumption that knowledge requires knowledge-independent justification, particularly after certain versions of externalism like reliabilism became popular. We shouldn't forget that before Goldman created process reliabilism, he preferred claiming that forgotten evidence cases were examples of knowledge without knowledge-independent independent justification, and gave an analysis of knowledge that included no justification condition at all.[56]

The view proposed so far as a second-order view about epistemic justification. It is no objection to this view that reasons talk is superfluous in a good account of knowledge in some cases if exactly the same claim can be made about justification talk. And I do not think one can simply revamp the objection by replacing "knowledge" with "justification" in the earlier paragraph. For it is just as plausible to claim that recollection *qua* ability provides justification as it is to say that it provides reasons for belief. As far as I can see, one is superfluous iff the other is.

A different version of the objection is that some promising theories of epistemic justification make no appeal to reasons, or are not about reasons. In the obvious cases, these theories will be externalist theories like process reliabilism. Consider the classic process reliabilist's account of noninferential justification: S's belief that P is *prima facie* noninferentially justified iff it is the output of a reliable belief-independent belief-forming process, and *ultima facie* noninferentially justified iff it is *prima facie* noninferentially justified and there is no alternative reliable process available to S which would have led S not to believe that P. Reasons-talk appears nowhere in the *analysans*. Wouldn't the truth of this theory therefore undermine some of my proposals?

No. This objection again confuses two levels of theorizing. One question we can ask is how to understand the supervenience between *epistemically normative facts* and the *nonnormative facts*. A theory that answers this question is a substantive, first-order theory of epistemic justification. It is a constraint on a kind of this theory that no normative concepts—at least none in the same family—

---

[56]See Goldman (1967: 369-72). This kind of view would play nicely into my own hands. If we had such a view, we could claim that knowledge is what gives epistemic reasons without also accepting the knowledge-first view. We could, that is, reverse the order of analysis between knowledge and justification without *also* supposing that knowledge is unanalyzable. I suspect this is the best view, as I indicated in other footnotes.

appear in the *analysans*. A different question we can ask is how *one* kind of normative fact—viz., S's being propositionally epistemically justified—might reduce to a different kind of normative fact—viz., S's having sufficient reasons for belief. A theory that tries to answer this question is a second-order, metanormative theory. Process reliabilism is not a theory of this kind. It hardly follows that it is incompatible with such a theory, just as it is clearly false that embracing, say, hedonism commits one to the denial of the buck-passing account of value.

These are different kinds of theories. You might think the second kind is uninteresting. I can't say that I agree. As I noted earlier, there is hot dispute in metaethics about the relations between different normative properties in the broad sense—whether, for instance, evaluative properties like *goodness* reduce to deontic properties like *having properties that give sufficient reasons*, as buck-passers like T. M. Scanlon would insist, or whether the reverse is true, as buck-stoppers like Roger Crisp insist, or whether *both* notions are fundamental. Part of the reason I'm interested in claims like PEJ=SER is that I'm fond of the broad view—one that crosscuts ethics and epistemology—that the fundamental normative thing is the *reason-for* relation. I pass the buck on value. It would disturb me if I couldn't reduce facts about justification to facts about reasons.

We can now see how the second version of the objection is problematic in two ways. Firstly, the reasoning behind it would overgeneralize and collapse a distinction very familiar in (meta-)ethics—e.g., the difference between metanormative theories of value like the buck-passing theory, and substantive, first-order theories of value like hedonism, as well as between metanormative views of moral permissibility that view it as the having of sufficient moral reasons, and first-order theories of moral permissibility like act utilitarianism.[57] Secondly, the datum with which it starts could establish *nothing* about the status of a second-order theory. *Anyone* who thinks that epistemic normativity supervenes on nonnormative facts will have to agree that there is *some* first-order account of epistemically normative properties in nonnormative terms. Of course, the theory might not be *systematic*, if something akin to particularism held in epistemology. But there would still be enough to get the same kind of argument going.[58]

Still, could process reliabilism, understood as a theory of epistemic justification, really *fit* with claims like PEJ=SER even if it is not strictly inconsistent with it?

On the face of it, it is completely unclear why it couldn't. Just take the process reliabilist to be claiming that facts about process reliability are the background conditions that are required for certain further considerations to be good epistemic reasons. To be sure, we would need to supplement this with a view about *what* the given reasons are. Once again, I'd suggest a view on which they are facts or propositions. In the basic cases, this would turn heavily on the suggestions I made in replying to the objection in §2.5. Facts about the reliability of processes would be what confer the status of being reasons onto these facts or propositions. *Structurally*, there is no difference whatsoever between this claim and what one would claim if one were an experiential foundationalist and accepted the reasons-first view. The experiential foundationalist just has a different view about what the relevant background conditions are, and what it takes for epistemic reasons to be sufficient. He and the process reliabilist can agree about what the reasons are, and agree that epistemic justification turns on having sufficient reasons. They need only disagree about what *makes* the reasons good, and what it takes for them to be *sufficient*.

---

[57]While the intuitive distance between permissibility/obligation and reasons is smaller than the distance between value and reasons, buck-passing accounts of permissibility/obligation are certainly nontrivial. Cf. Bedke (2011) for a nice discussion. It's worth noting that Parfit passes the buck on value but not on obligation: he thinks it's at least intelligible that there is a basic property of *just-plain-forbiddenness*.

[58]Cf. McKeever and Ridge (2006) and Strandberg (2008) on supervenience and particularism.

2.9.   *Having reasons (again)*

There is a final worry one might have, specifically about the account of propositional epistemic justification. It stems from the apparent fact that the having of reasons can only be understood in epistemic terms. If these terms are also paradigmatically *normative,* Reasons Basicness would seem to fail. For there would then have to be further normative features *not* analyzable in terms of reasons *via* which we'd have to understand having reasons.

Why think having reasons will need to be understood in epistemic terms? Short story: this is just the most natural thing to say if we generalize from simple observations in the practical case. Go back to our earlier case where there was, completely unbeknownst to you, a lion that would eat you behind Door A. We wanted to say that the lion's being behind Door A was a reason not to open it, but that you did not have this reason. Why not? The answer seems obviously to turn on "completely unbeknownst to you" bit. You didn't have the reason because you *weren't in a position to know it*. If this is how we understand having reasons in the practical case and we want high-level symmetry, we'll say the same about having epistemic reasons.

Suppose we just take the appearances here at face value. Do we face a problem? My honest reply is that we simply do not. But this is part of a systematic view that would take a whole further paper to defend. The defense is ultimately not needed, since there are other ways out. But I might as well be honest about what I take to be the correct way to avoid the objection.

My view is this. Williamson is partly right: knowledge is prior to justification. But the Goldman of the late 1960s was also right. Knowledge is analyzable, and *not in terms of justification*.[59] On this latter score, I would accept not the young Goldman's causal theory of knowing, but rather an explicitly nonnormative variant of what Chris Kelp (ms) calls the "safe-apt view": knowing is true believing that manifests a reliable cognitive character trait or ability, where the believing is also basis-relative safe. Properties like reliability, character traits and abilities are not normative properties. Of course, one could beef things up and use "virtue-theoretic" *terminology* to state this view. That talk is normative in *some* sense—though I'd say a *mere constitutive* sense, like that associated with "This is a good knife". But one does not need to use such talk. And the properties that actually do the work in the view are clearly nonnormative properties.[60] If we can *directly* ground facts about knowledge in nonnormative facts, we can clearly appeal to knowledge in analyzing the having of reasons without circularity.[61]

With this background view in place, here is my view. Good epistemic reasons are facts, very often worldly facts. To have these reasons, one must *at least* be in a position to know that they obtain. Being in a position to know that they obtain simply involves being in a position to believe truly that they obtain in a way that manifests some reliable ability, and that is basis-relative safe.

---

[59]There has been some interest in reviving this idea; see, e.g., Kornblith (2008). And it is worth adding that the idea wasn't just in Goldman (1967) in the early post-Gettier literature. Armstrong (1973) defended the idea. Dretske has also embraced in many different places over his career. See esp. Dretske (1981: ch.4).

[60]Some might say that abilities are normative properties. Some virtue epistemologists—e.g., Greco (2010)—do. But this is wrong. Abilities are just powers directed at certain manifestation states. Sometimes the manifestation states might *independently* be good/bad, or be ones we have reasons to bring about or eliminate. But their independent goodness/badness would not make the *abilities* normative. (Compare poisonousness. I take it that poisoning is *ceteris paribus* a bad thing. That does not make poisonousness is a normative property.) If we talk explicitly about "competences", we may be discussing *constitutively* normative properties—ones that give rise to *attributive value* facts, like facts about *good cognition*. I doubt these are genuinely normative. But even if they were, I can simply spin the view with abilities instead.

[61]I should emphasize that I don't doubt that facts about epistemic reason-relations themselves are *also* reducible to nonnormative facts. The point is simply about the *order* and *directness* of the reduction.

Justification is then analyzed (partly) in terms of possessed sufficient epistemic reasons. Yes, knowledge is then needed to analyze justification. But that is simply the truth! Call this the *Weak Knowledge-First View*. We get this reduction tree ('←' means 'partly reduces to'):

propositional justification	←	having sufficient reasons
having reasons	←	being in a position to know these reasons[62]
being in a position to know some reasons	←	being in a position to believe truly that these reasons obtain in a way that manifests a reliable cognitive character trait or ability, and that is basis-relative safe

But one does not need to accept my Weak Knowledge-First View or the analysis of knowledge I favor to avoid the worry. One could make different moves, and still keep Reasons Basicness.

A natural alternative is to put mental states stronger than belief but weaker than knowledge first. One could claim that, in the foundational case, the way we come to have reasons is partly by being in these mental states. As long as these states do not clearly involve normative properties—e.g., they're not states like *justifiedly believing that P*—we won't contravene Reasons Basicness.

To see the version closest in *upshot* to my view, we can start with perception. Perceptual states clearly *sometimes* make it the case that we have their contents as sufficient reasons. When I see that there is a red cube, that makes me have a sufficient reason to believe that there is a red cube, other things equal. But how does my perceptual state do it? I doubt that it is simply by involving a veridical perceptual experience *with the content that there is a red cube*. After all, consider speckled hens. My visual experience could represent a hen with 77 speckles on its facing side. I do not *have* a reason to believe that there is something with at least 77 speckles before me *simply by having this experience*. Reflecting on cases like this suggests that selecting the contents of veridical experience that are the reasons we get to *have via* experience is a task one is not going to be able to complete without some kind of reliable ability constraint—though one might explain the *providing* of reasons simply by appeal to the experience (given the distinction between reasons provided for one and reasons had by one from §2.1). The constraint might be:

> it's perceptually seeming to one that P makes one have P as a reason only if it's perceptually seeming to one that P manifests a reliable perceptual ability of one's *vis-a-vis* P-like contents.

This could be a subitizing ability, if we are dealing with the seeing of speckles. In good cases, this constraint will plausibly make the difference between *seeing that P*, where P is a content of the experience (i.e., one of its accuracy conditions), and *experiencing the fact that P* in a *de re* fashion.[63]

---

[62]I doubt this will be the *whole* story. As we saw earlier in §2.1, we must take care to distinguish between a reason being *provided* for one to believe Q, and one's having that reason to believe Q. Having a reason does require that one be provided with it. Knowledge will help to provide reasons, and so will help to explain having reasons. But it may not do all the work. Again, take the inferential case. You know P. P entails Q, but by an inference you can't see. P is itself a sufficient reason for you to believe Q. But you do not have sufficient reason to believe Q. Adding more knowledge may help in *some* cases. But if the inferential link is suitably basic, and you don't understand it because you are like Carroll's tortoise, what is really needed is going to be a deeper reasoning ability. None of this creates further trouble of the kind I'm addressing in this section, since all the extra detail will, I believe, be directly analyzable in nonnormative terms.

[63]There clearly is such a thing as a *de re* experience of a fact. After all, one can uncontroversially see an *event* without having the concepts required to isolate the elements of that event. Your child might see a great performance of Beethoven's 9th Symphony. But events just are *facts plus times*, if we hold the familiar view of events as property exemplifications at times defended by Kim (1976) and the familiar view of facts as property exemplifications *simpliciter* defended, e.g., by Armstrong (1997).

The point extends to memory, since there is a distinction between experientially remembering, in a *de re* way, certain facts specifiable in propositions P, Q, R, etc., in one's past, and *remembering that* P, Q, R, etc. An analogous point holds for introspection too, since there is a distinction between being introspectively aware in a *de re* way of having a mental image with a certain content C, and introspectively seeing that one is having a mental image with C. Indeed, an analogous point holds even for intuitive seeing. A tired mathematician on the brink of a discovery might *sense*—at the tip of his intellectual tongue *via* some reliable metacognitive signal—the solution to a certain problem, but be too tired to follow through and intuitively see *that the problem is soluble*. Or consider a more familiar case. Late at night one is thinking about some philosophical question, and senses that there is an answer, and even senses the details, but is unable to put them together simply out of exhaustion. The details may be there, just like with the speckled hen. But one may not fuse them in the way required for intuitively seeing the answer. This may simply be because one lacks the attentional resources at this point in the evening.

Bearing all this in mind, one could adopt the following alternative view. In foundational cases, we come to have reasons in virtue of being in a position to enter into factive mental states such as seeing that P, remembering that Q, introspecting that R, intuitively grasping that S, and so on. In saying this without also accepting the Weak Knowledge-First View, one rejects Williamson's view that knowledge is the most general factive mental state. One simply claims that these factive mental states are more fundamental, and are not *determinates* of knowledge. They *provide* knowledge, rather than *fall under* it. Call this the *Non-Epistemic Factive View*. It is McDowell's view.

Of course, an explanation is going to be needed of why these states can make it the case that one *has* their contents even when closely related states—such as seeing *de re* the fact that P, remembering *de re* the fact that Q, introspecting a mental image of the possible fact that R, and intuitively seeing *de re* the solution to some problem while exhausted—cannot make it the case that one has their contents as reasons. The explanation will plausibly invoke a reliable ability constraint of some kind. And I suspect that Dretske's classic work on *seeing that* will generalize to suggest a counterfactual constraint as well.[64] Ultimately, one will end up with the spirit of my own view. Seeing that P, remembering that P, and so on, simply are manifestations of specific reliable abilities on a subject's behalf, and the representational outputs in which these seeings, rememberings, etc., consist will have to be safe. That is what they have in common. It is also the best general explanation of why they are capable of making it the case that we have their contents as reasons. As long as one rejects Williamson's view that knowledge is the most general factive mental state, one does not end up committed to my view.

If one doesn't like views that stress factivity, responding to the original objection is *even easier*. Consider Pyror (2000)'s epistemology of perception, and Huemer (2001, 2007)'s phenomenal conservatism. These views replace our factive states with *non-doxastic seemings*. These theorists may have it that one has reasons in foundational cases only if (and often if) they are the contents of seemings. Call this the *Non-Epistemic Non-Factive* View. Now, I doubt whether this view works without reliability constraints. One needs to restrict the contents of the states one might reasonably call "seemings" (e.g., perceptual experiences, in Pyror's case). The problem of the speckled hen is a key reason for this. While there are ways out, they create "problems of scatter": the restrictions that one ends up endorsing on contents eligible to be possessed as reasons will call for explanation.[65] Without an explanation in terms of something like reliability, the restrictions

---

[64]See Dretske (1969). The reasons for distinguishing safety constraints from his preferred counterfactual clauses (which are sensitivity constraints) obviously predate this classic work. One wonders how he would respond to these reasons now. At any rate, safety-based variants of his view are obviously possible.

[65]The use of this "scatter" talk in epistemology comes, as far as I know, from Sosa (1991: 128): "A third problem

look arbitrary ("scattered"). Still, phenomenal conservatism is a partial account of how we get to have certain propositions as reasons. If it doesn't collapse into a variant of the account most recently considered, it will be a further account framable in epistemic terms.

So, there are many ways to understand the having of epistemic reasons that do not require the *having* relation to be cashed out in *normative* terms. We can put them on a spectrum:

← **High Bar**   **Low Bar** →

Williamson's View	Weak Knowledge-	Non-Epistemic	Non-Epistemic
	First View	Factive View	Non-Factive View

Views that set even lower bars are conceivable. Schroeder (2011), for instance, has argued that all it takes to have a reason is to *believe* that it obtains.

With this rich spectrum of possibilities, I cannot see that one can reasonably complain that the only plausible account of the *having* relation will require us to reject Reasons Basicness.

3.   *Criteria of quality for epistemic reasons and theories of epistemic value*

Far from alienating externalists, viewing epistemic justification as turning on having sufficient possessed epistemic reasons may give them a significant advantage. To see the deepest reason why, we can start to reflect on the need for *criteria of quality* for epistemic reasons—criteria that tell us *when* epistemic reasons are *strong* reasons, or are *stronger than* the competition. When we turn to why there are problems of deontic significance for epistemic rationality in Chapter 3, we'll see even more clearly how here putting reasons first creates an externalist advantage.

How do we get such criteria? A fruitful strategy is to try to illuminate the criteria of quality for epistemic reasons by appeal to high-level intuitions about *epistemic value*, together with independently plausible general assumptions about the relation between value and reasons. It certainly seems like there ought to be a story to tell of this form. After all, when certain epistemic reasons for believing P are *sufficient*, they are *better* than any epistemic reasons against believing P. Whether certain epistemic reasons for belief are *better* than others will surely have *something* to do with what we ultimately take to be *epistemically good* or *bad*. Of course, this is highly schematic; I'll turn in a moment to consider some more concrete ways of developing it.

Now, as a defender of Reasons Basicness, I cannot think that facts about epistemic value are metaphysically fundamental within the epistemically normative domain. But this doesn't prevent me from giving a different kind of role to epistemic value: it could play a fundamental *conceptual* role. Structuring our *beliefs* about what reasons are strong or weak by appealing first to our *beliefs* about epistemic value does not conflict with Reasons Basicness. Reasons Basicness is a claim about *metaphysical* structure within normative domains. It is not incompatible with allowing intuitions about value to drive the *first-order* task of specifying *when* reasons are strong or weak. A direct analogy with the practical case makes this obvious. No one should think that a defender of a buck-passing account of value cannot be an act consequentialist. Yet act consequentialism turns exactly on the plausible thought that the act we have *most reason to do* has to be the act that *produces the best outcome* or that *makes things go best*. This is a high-level thought about the relationship between

---

with evidentialism is related to the second. John Stuart Mill would often object to 'intuitionist' moral theories by deploring their lists of retail intuitions with no apparent unity, thus charging them with a problem of 'scatter'. Suppose evidentialism stops with a scattered set of principles. For example, suppose it includes principles admitting green and blue but ruling out bleen and grue, and suppose it includes principles that allow direct introspection of triangularity but not direct introspection of octagonality or of 23-sidedness. Obviously there would then be a problem of scatter."

what we have most reason to do and what outcomes are best. The appeal of this thought is a big part of what drives attraction to act consequentialism. Yet act consequentialism is a first-order theory, and so is certainly compatible with buck-passing.

So, even for a friend of Reasons Basicness like myself, it is worth considering this approach to questions about the strength of epistemic reasons, which I will call **value-driven methodism**.

3.1. *Teleological criteria of quality*

Strikingly, only one general way of pursuing value-driven methodism has ever been considered: **strong epistemic teleology**. Strong epistemic teleology should be understood as *adding* to a purely axiological claim I'll call **weak epistemic teleology**. This weaker claim says that:

> all intrinsic epistemic value is *final* epistemic value—epistemic value to be *produced as a goal* or *end*—and all extrinsic epistemic value is *instrumental* epistemic value *relative to these goals/ends*,

and that

> all intrinsic epistemic disvalue is final epistemic disvalue—epistemic disvalue to be *eliminated* and *avoided* as an end—and all extrinsic epistemic disvalue is instrumental disvalue.[66]

These high-level claims would be coupled with a substantive account of intrinsic epistemic value and disvalue, typically one on which the relevant final values and disvalues are *states* such as true belief and false belief (or knowledge and ignorance). While most often simply *presupposed* in one form or another, weak epistemic teleology is nontrivial. Not all think the intrinsic/extrinsic distinction is the same as the final/instrumental distinction.[67] I'll argue in Chapters 4 and 5 that this assumption is *crucially* false for epistemology. For now, I'll assume it for illustration's sake.

Strong epistemic teleology combines these axiological claims with a background assumption about the relation between the (epistemically) deontic and the (epistemically) evaluative. A weak version of the assumption that would get us strong epistemic teleology is

> *D-E Supervenience*: Epistemically deontic facts—e.g., facts about whether one has sufficient or decisive reasons to (dis)believe P—supervene on epistemically evaluative facts, so that a change in epistemically deontic facts requires a change in epistemically evaluative facts,

where "epistemically evaluative facts" refers not just to *general* facts like the (putative) fact that *true belief is finally epistemically valuable*, but also somewhat more particular facts, like the fact that holding *the rule of believing for R-type reasons in C-type cases would be an optimific epistemic rule*.

The most obvious plausible example of strong epistemic teleology is **Epistemic Rule Consequentialism**. On this view, believing P for a reason R in a case C would be epistemically right (and so R would be sufficient) iff complying with the rule of believing P-like claims on R-like grounds in C-type cases would be *epistemically optimific*—i.e., it would produce a higher ratio of epistemic value to epistemic disvalue than any other relevant alternative rule.[68] If the slogan for

---

[66]Scanlon (1998: 80) nicely expressed the more general view with these words: "To be (intrinsically) valuable is to be to be promoted." Though "produce" is, I think, a slightly more accurate word than "promote".

[67]The idea is questioned *as such* in Korsgaard (1983), Rabinowicz and Rønnow-Rasmussen (1999) and Tannenbaum (2008). Rejection of the idea is also naturally encouraged by many other theorists' views. Scanlon (1998), for instance, would surely reject it. So would Anderson (1995). Indeed, I think many—e.g., Foot, Thomson, and Geach—who have been skeptical about the idea that it makes sense to regard *states of affairs* as good or bad and who emphasize the attributive nature of 'good' would reject this claim.

[68]The view could also take a satisficing form rather than a maximizing form, as some friends of this view would

ethical rule consequentialism is "What if everyone did that?", the slogan for Epistemic Rule Consequentialism might be "What if we always reasoned like that?", where "reasoned" is used broadly to mean "formed (doxastic) attitudes for reasons". A species of this theory is *Veritist Rule Consequentialism*: believing P for a certain reason R in a case C would be right (and that reason would be *sufficient*) only if complying with the rule of believing P-like claims on R-like grounds in C-type cases leads to a high ratio of true to false beliefs. Relative to a veritist version of weak epistemic teleology, this requires D-E Supervenience. After all, relative to such a view, it follows that if you fix the facts about what cases of believing for reasons comply with which general rules, and also fix which rules are epistemically optimific, you thereby fix whether believing for a certain reason would count as believing for a sufficient reason in any case.

One finds this kind of strong epistemic teleology in Goldman's work, though his terminology has varied over the years. Initially, in his (1980), he did not use 'value' talk at all but rather talked about *goals*, and focused on the *rightness* of "doxastic decision principles" ("DDPs"). He wrote:

> The choice of a DDP clearly depends on the *goals* of cognition, or doxastic attitude formation. A very plausible set of goals are the oft-cited aims of *believing the truth*—as much truth as possible— and *avoiding error* […]. [L]et us proceed on the assumption that *some* combination of true belief and error avoidance is what we seek in a DDP.
>   Given the aim of true belief and error avoidance, the right DDP is apparently the one that would produce optimal results in terms of true belief and error avoidance. It is the DDP that would have such optimal results in the long run for the sum-total of cognizers.[69]

Combining the objective teleological criterion here with Goldman's later ideas, we get a view that clearly counts as a kind of strong teleology. In his (1986), Goldman analyzed epistemic justification in terms of what would be permitted by the set of right J-rules, where these were like the DDPs considered in his (1980). In his (1999), Goldman adopted value talk and defended *veritism*: the view that the only intrinsic value in the epistemic domain is *true belief* (and the only intrinsic disvalue is false belief). If we put together the ideas from these different phases of his thought, we get strong teleology: facts about whether a belief is justified supervene on facts about what's epistemic valuable together with nonnormative facts (e.g., how everyone following some DDP or J-rule would in the long run lead to a high ratio of true to false beliefs).

Strong epistemic teleology has substantial *prima facie* appeal. It has the same kind of appeal that consequentialism in ethics has when considered abstractly. For any property that the strong epistemic teleologist will regard as genuinely deontic, there is an *extremely simple* story about why we should care about having doxastic attitudes with that property: having them is *actually generally conducive to realization of the goals of cognition* and hence, assuming weak epistemic teleology, to what is epistemically best. How could one want anything else? The same rhetorical question has force when one considers consequentialism in ethics. What sense could there be in licensing acts that *systematically make things go worse*, or less well than they could have gone if some alternative had been pursued? Even nonconsequentialists appreciate the force of this thought:

> [I]t is hard to see how it could be perfectly all right to fail to do what would make the world go better than it otherwise would. If you really would make the world go worse if you did a thing, then are you not called on morally to refrain from doing it? What else is there for the right to turn on than how good the world will be if you act in this way rather than that?[70]

---

be eager to note. But since none of my own objections to Epistemic Rule Consequentialism or teleology in general will turn on whether the view takes a satisficing or maximizing form, I focus on the maximizing form for simplicity's sake.

[69]Goldman (1980: 32).

[70]Thomson (1997: 274).

Crucially, though, the analogue of this thought has no clear force in motivating strong epistemic teleology *unless* one assumes weak epistemic teleology. Ultimately, I'll want to question it.

For now, I'll simply note how standard the assumption is. Almost all epistemologists think that there are such things as epistemic goals.[71] Many would simply identify epistemic values with epistemic goals. Having done this, the questions for epistemic value theory are *then* always taken to be about about *which* things are final goals/intrinsic values. Can everything be explained by appeal to true (*false*) belief as the fundamental epistemic value (*disvalue*)? Or must we also accept knowledge (*ignorance*) as a fundamental epistemic value (*disvalue*)? Or understanding (*confusion*)?

Once weak teleology is accepted, strong teleology can feel inevitable. How could it be epistemically right to believe for reasons of a certain kind, when beliefs formed for reasons of that kind are *systematically guaranteed to be objectively likely to be intrinsically epistemically bad beliefs* (false ones)? How could a reason for believing be sufficiently good if, in general, beliefs held for reasons of that sort would be objectively likely to be epistemically bad beliefs (say, false ones)? These rhetorical questions are forceful. Their force suggests that the most natural view to take about the criteria of quality for epistemic reasons will be an externalist one: whether believing for a certain reason would be actually objectively likely to be intrinsically bad believing is not typically something one can discover by reflection alone, and is certainly not a fact about one's internal states. This is particularly obvious if the intrinsic disvalue is false belief or ignorance.[72]

We'll be returning to this kind of motivation for an even more far-reaching kind of externalism in Chapter 3. Strong epistemic teleology affords the most direct route to skepticism about the deontic significance of rationality, once the correct views about rationality are seen in Chapter 2.

3.2. *Strong anti-teleology and criteria*

Now, I have never seen general explicit alternatives to weak or strong epistemic teleology. I'll briefly sketch some alternatives in this subsection and the next to preview later ideas, and also to set up a point I'll make near the very end of the chapter about how deontic and hypological properties in epistemology could be linked. I am not going to try to *argue* for any alternative now. That task will come *much* later, only after we've seen where epistemic teleology leads us.

One way to get an alternative is to reject weak epistemic teleology. I'll call any view that rejects this claim a kind of *strong anti-teleology*. To reject weak epistemic teleology, one must argue that:

> not all intrinsic epistemic (dis)value is final epistemic (dis)value (i.e., (dis)value to be produced(/eliminated) as an end or goal), and not all extrinsic epistemic (dis)value is instrumental epistemic (dis)value.

To grasp what some alternatives would look like, it is crucial to look at the ethical case first.

One way to oppose weak teleology in ethics is to note that there are intrinsic values that are not best understood in terms of the fact that producing them is what we ought to do. *Persons*, for instance, might be thought to have a special kind of value—viz., *worth* (or "*Kantian dignity*").

---

[71] See Berker (ms) for a fine overview of exactly how common this assumption is (and some arguments against strong teleology, which I shall discuss critically in Chapter 4).

[72] Of course, if one tried to adopt the view that irrationality is intrinsically epistemically bad, one might be able to avert this conclusion by then arguing for an internalist view of (ir)rationality. But it is, I think, extremely plausible that rationality and irrationality are at best extrinsically epistemically valuable/disvaluable. And I think virtually everyone in the literature on epistemic value—even the people antecedently disposed to be internalists (cf. BonJour (1980))—would agree about this.

Persons *call for respect*, and the fact that they call for this distinctive pro-response *just is* what it is for persons to have the special kind of value that is their worth, on a buck-passing account that I'd strongly favor. All the same, we clearly do not have powerful and non-optional reasons to see to it that there are *more people* in the world, even more happy people. The worth of persons is clearly a non-teleological kind of value. As Jan Narveson famously put it: while we ought to make people happy, we have no obligation to make happy people.

One might go on to say—plausibly—that the distinctive value that persons have is intrinsic. It is not *derived* from the fact that persons are *mere containers of pleasurable experiences*, as a crude hedonistic total utilitarianism would wrongly suggest. Indeed, one might plausibly suggest *exactly the reverse*, and say that the value of increasing total welfare *rests on* the prior *worth* that persons have. On this view, the deepest intrinsic value for morality to attend to is not a kind of final value—not a kind of value to be produced as an *goal*. Of course, Kantians may like to use the terminology of persons being "ends in themselves". But Narveson's dictum suggests that this is mostly just nifty rhetoric: there is no moral duty at all to create more (happy) people.

It is clear enough how this view implies that not all intrinsic value is final value, where final value is value to be produced as an end. Notice that this view *also* implies that not all extrinsic value is instrumental. Suppose that one succeeds in having respect for persons. That very act of having respect clearly has value. But the value is, I think, most plausibly viewed as being *derived* from the value of its object—i.e., from the worth of the person. Acts of respect are only good *because* the persons at which they are directed antecedently have worth—i.e., call for respect. Clearly, though, this value derivation is not *instrumental* derivation: showing respect for persons is ideally not a means to anything. It is just done, *even though* its value is derivative from the worth of persons. This suggests, compellingly to my mind, that extrinsic value needn't be instrumental.

There is a precise analogue of this view in epistemic value theory, though, remarkably, it has never been explored as such, and never comes up in the vast recent literature on epistemic value. On this view, truth (understood as a property of doxastic representations) may be the only intrinsic value in the epistemic domain, but its value does not consist in the fact that we have some general reason to produce more true beliefs and eliminate false beliefs. Instead, the value of truth simply consists in the fact that we ought to give it respect as the standard for belief-formation. Having respect for truth as the standard for belief formation could amount to manifesting *full commitment in every case to the fundamental rule to believe P only if P*. What, in turn, does it take to manifest full commitment to this rule? At the very least, it takes complying with all the requirements of epistemic rationality. That, at any rate, will be my ultimate view.

I'll call the general view of which it is a particular case **epistemic Kantianism**. This view is most at home in a veritist setting. So ensconced, amounts to this claim:

> The intrinsic epistemic value of truth—understood as a property of some doxastic representations—consists in the fact that we ought to epistemically respect it as the standard of belief formation for its own sake.

One may add the following further claim to get a more ambitious view (which I'll ultimately do):

> **Kantian veritism**: All other epistemically valuable states derive their value from the value of truth—understood as a property of doxastic representations—by, in one way or another, being suitably related to the fact that we ought to respect truth as the standard for doxastic attitude-formation.

The conjunction of epistemic Kantianism and Kantian veritism is deeply non-teleological. It is

compatible with this view that a *given state* of believing truly may have no value, and indeed that there may be nothing epistemically good about a vast body of true beliefs. If these states were all arrived at by a sheer reckless cognitive accident, there need be nothing epistemically good about this believer's believing truly, or even having a completely accurate huge system of beliefs. Of course, in ordinary cases, many true beliefs *will* be obtained out of respect for the truth. It is *hard* to be so cognitively reckless without simply ceasing to be a believer. So, in ordinary cases, many states of believing truly will have some epistemic value. But it will be derivative.

As I'll argue in Chapter 4, this view has the resources to straightforwardly explain the distinctive extra epistemic value of knowledge in veritist terms. One does not need to retreat to a non-veritist axiology to explain why knowledge is distinctively epistemically valuable. Truth can remain the fundamental thing in epistemic value theory. One need only reject a teleological picture of the value of truth. Knowledge may be distinctively epistemically valuable, and indeed be generally more epistemically valuable than true belief, *because* when someone knows, they manifest respect for truth in believing. Since true beliefs can be gotten without respect for truth, they may be *worthless*—worthless, it's worth noting, in terms that a veritist can perfectly well grasp, as long as she is not also a teleologist.

Notice, finally, that there is a way in which epistemic Kantianism may *extensionally subsume* Epistemic Rule Consequentialism. It might turn out that *as a matter of fact* we can only have full respect for truth as a standard of belief by following the rules that lead to the highest ratio of true to false beliefs. And it might also turn out that we have to do *more* to have full respect for truth than simply use the optimific rules. I suspect both of these things are true. If so, epistemic Kantianism could very often agree with epistemic rule consequentialism about what we ought to believe/disbelieve, and simply disagree about *why* we ought to believe/disbelieve certain things in those cases. The fundamental explanation of why we ought to believe/disbelieve is to be given in respect-based, non-teleological terms.

This observation is structurally related to the core theme of Derek Parfit's *On What Matters*—the theme that explains why this book was once intended to be called *Climbing the Mountain*. Parfit's view is in effect that Kantianism and rule consequentialism make virtually the same predictions when both are formulated in the most plausible terms. Hence the mountain-climbing metaphor: these theories are climbing to the top of the same mountain, just from different sides. Of course, this crucially leaves it open which of these theories provides the better *deep* explanation of why we ought to act in certain ways. This is one lingering source of dissatisfaction about Parfit's project: because the theories disagree about the explanation, *even if* they agree on all cases, that is hardly the end of ethical understanding.[73] Something similar *may* go for epistemology, at least when Epistemic Rule Consequentialism is formulated in a certain plausible way. I will be investigating the possibility of this kind of Parfitian convergence in later chapters.

3.3. *Stronger anti-teleology?*

Now, strong anti-teleology of the kind envisaged in the last subsection attacks only weak epistemic teleology. It does not by itself attack D-E Supervenience. The epistemic Kantian simply has a different picture of what it is for truth—understood as a property of doxastic representations—to be epistemically valuable. It is open to her to allow that the epistemically deontic facts supervene on the epistemically evaluative facts. She simply overhauls our understanding of *what* these evaluative facts are.

But D-E Supervenience is not a compulsory assumption. It is structurally similar to an assumption

---

[73]Indeed, that point can be turned into a concern about Parfit's convergence thesis, as Otsuka (2009) claims.

in ethics that is deeply controversial. Deontologists in ethics standardly deny that if one fixes the evaluative facts, one thereby fixes the deontic facts. Many would claim that if an act would produce the best outcome, or even would comply with an optimific rule, that does not answer the question of whether one would be permitted to do it. One must also know whether one would violate a restriction in doing it, a restriction that may not coincide with any optimific rule. Strong epistemic Kantians could make similar further claims and reject D-E Supervenience. This would make them, for lack of a better term, "stronger epistemic Kantians".

It might seem *best* to respect the truth at all times, and not just at some times. So, if we're considering the value of outcomes of possible doxastic "decisions", understood broadly enough to include being in the doxastic state itself, it isn't clear that epistemic Kantians must accept D-E Supervenience. In reflecting on this, we find an analogy with one version of the "paradox of deontology". Deontological restrictions apply even when violating them would entail a greater number of violations *of the same restriction*: indeed, this applies even when *you* would be the one bringing about the greater number of violations *at some different time*. In a paper in which he suggests that "agent-centered" is the wrong modifier for restrictions, Richard Brook asks us to consider the following cases, which I reproduce *verbatim* along with his remarks:

> You are at the zoo with two children who are making a scene. Becoming angry, you toss them into the lion's den. Horrified, you come to your senses and notice that they can only be saved if you toss a third child (who just toddled along) into the back of the den. The beast would then be distracted and you could quickly leap in and save the first two. What should you do? Whatever you do, you have initiated a chain of events that results in a child's death. If you do nothing, you have intentionally killed the first two children. If you sacrifice the third child, you kill it in using it as a mere means to save the others. Or consider this variation of Darwall's example about betrayal. Captured by the enemy and under some pressure, I reveal the hiding place of two comrades. They will be killed if discovered. I could prevent this by revealing the whereabouts of a third comrade, perhaps of more importance to the enemy. Whether or not I reveal her whereabouts, I have intentionally betrayed someone. And equally important, if I fail to betray the third comrade *I still remain directly involved* in the betrayal of the first two.
>
> Yet it seems clearly wrong to kill the third child or betray the friend; as wrong, in fact, as in any of the standard cases where you might harm one person to aid others.[74]

The same kind of structure shows up in the cognitive domain, and strongly suggests that epistemic Kantians ought to deny D-E Supervenience. Here I foreshadow a later discussion of Selim Berker's work. Suppose that unless you believe a false proposition now, you will end up believing many clearly false propositions in the future, which you would otherwise not end up believing. That gives you no epistemic reason now to believe the clearly false proposition. In the same way in which the "separateness of persons" calls for independent, person-by-person applications of the restriction (as in Brook's cases), so, Berker plausibly claims, the "separateness of propositions" calls for independent, proposition-by-proposition applications of the truth-respecting constraint. Even in observing this, one could keep an axiology on which it would be epistemically better for you to believe the false proposition now, and thereby violate the truth-respecting constraint now. If so, D-E Supervenience will obviously fail.

This is a way in which one might end up having to reject D-E Supervenience even if one accepts strong epistemic Kantianism. *Stronger* epistemic Kantianism holds that the response that is called for by the value of truth is *respect*, and adds that the response is called for on a *proposition-by-proposition* basis, in much the same way as the dignity of persons calls for respect on a *person-by-person* basis. While I don't find the analogous version of Kantian ethics compelling, stronger epistemic Kantianism is, I think, an attractive view. Much more will be said about it later.

---

[74]Brook (1991: 197–8).

4.   *Hypological properties and separability theses*

I turn to a final set of preliminary distinctions and points. As I noted at the outset, it is commonplace in ethics to distinguish properties linked with a person's criticizability and appraisability from properties like obligatoriness, permissibility, justifiedness, rightness, and support by reasons. Following Zimmerman, I called the former **hypological**. One difference between these properties and deontic ones is that what is targeted in the first instance is the *agent*, whereas what is obligatory, permissible, justified, or supported by reasons in the first instance are her *acts*. Hence the familiar "act/agent distinction" in ethics.

This isn't the only difference. A further difference can be seen by first noting a similarity. Hypological and deontic properties both have *valences* that can be aligned with one another:

	DEONTIC	**HYPOLOGICAL**
+	A-ing is required (= *decisive reasons for*)	S is praiseworthy in A-ing
	justified (= *strongly sufficient reasons for*)	S is positively assessable in A-ing
	permitted (= *sufficient reasons for*)	S is free from criticism in A-ing
−	A-ing is unjustified (= *sufficient reasons against*)	S is negatively assessable in A-ing
	impermissible (= *decisive reasons against*)	S is blameworthy in A-ing

There would be a tight link between the deontic and hypological if we could defend the

> **inseparability thesis**: Whenever S does an act (or has an attitude) with a certain deontic valence, S can be imputed a hypological property of the same valence. Moreover, whenever S can be imputed with a hypological property of some valence for an act, that act itself has a deontic property of the same valence.

But the inseparability thesis fails in the practical domain. An agent can do what is required or permitted for the wrong reasons, or in the light of clear but misleading evidence that so acting would be wrong, or with false beliefs entailing this. While the act is required or permitted, the agent is blameworthy or at least negatively assessable. Conversely, an agent can do what is unjustified or impermissible in the light of clear but misleading evidence that so acting would be right, and act for this reason. While the act would be impermissible or unjustified, the agent would be free from criticism, and perhaps positively assessable or praiseworthy in so acting.

It is fairly plausible that one can *freely combine* deontic and hypological properties: just make someone's evidence is sufficiently misleading, or her beliefs or reasons for acting sufficiently off base. I call this the **strong separability thesis**. It is endorsed by some people in ethics.[75] I cannot call it an *orthodox* view, but it's a view that is taken seriously. What *is* orthodox in ethics is the view that deontic and hypological properties *do not always* align, so that for any act A with a deontic property D of some positive or negative valence, an agent who does A cannot always be imputed a hypological property of the same valence in A-ing. I'll call this view the **weak separability thesis**. The difference is clear. The strong separability thesis implies that it is possible for an

[75]See Zimmerman (1997) for a particularly clear example. Thomson (1986) and (1992) definitely seems attracted to the idea.

agent to do an act with some valence while being having a hypological property *of the opposite valence.* Someone could deny this, but concede that one needn't always be positively assessable for doing the right thing, and that one needn't always be negatively assessable for doing the wrong thing: one could be *neutrally* assessable—neither praiseworthy nor blameworthy.

While some epistemologists have also toyed with separability theses, these parallel claims are far from being as widely accepted or even considered. Many epistemologists systematically conflate them. One example is BonJour, who wrote:

> My contention here is that the idea of avoiding […] irresponsibility, of being epistemically responsible in one's believings, is the core notion of epistemic justification.[76]

While I ultimately embrace something like BonJour's claim, it is to be *earned,* not *taken for granted.* This claim is structurally analogous to saying that the core notion of acting rightly is acting responsibly, or in a way that merits positive appraisal for the agent. This is a false claim. Anyone who agrees should find it *surprising* if epistemology is structurally different in such a deep way. While I think it *is,* this is a *very* nontrivial claim that calls for argument and explanation.

Now, there is a fairly theory-neutral reason for thinking that there is more of a connection between the deontic property that has most interested epistemologists and the corresponding hypological concept. Epistemologists have been interested in the kind of epistemic justification that one has when one knows. This is *doxastic* justification, not mere propositional justification. On the high-level metanormative account offered in §2.6, doxastic justification requires believing for a sufficient reason *in the strong sense.* As I suggested in §2.2, acting for a sufficient reason in *that* sense requires that one's reason for believing is in fact a good reason, *and* that it is one's reason *because it is a good reason.* Doxastic justification turns partly on the epistemic agent's motivating reasons for belief, and requires a certain connection between *why* these reasons are the agent's motivating reasons and the fact that these reasons are *actually good.* In the practical case, it is plausible to think that if someone acts for a sufficient reason in the strong sense, she is *ipso facto* praiseworthy. If this holds in epistemology, BonJour's claim is partly vindicated: being epistemically responsible in one's believings would be required for doxastic justification.

It would not be fully vindicated. What follows is that

      (a)   doxastic justification suffices for epistemic responsibility,

not that

      (b)   epistemic responsibility suffices for doxastic justification

or, equivalently, that

      (b*)   a lack of doxastic justification suffices for epistemic irresponsibility.

BonJour does, I think, believe (b) and (b*). It is much harder to defend these claims. (b) and (b*) are incompatible with the best versions of externalism, and not to a fault. Consider the so-called "new evil demon problem". This isn't exactly a problem, as I see it, but just an illustration of the possibility of a deontic/hypological separation. Your internal twin in the demon world does not

---

[76]BonJour (1985: 8). Also consider Ginet (1975: 28): "One is *justified* in being confident that *p* if and only if it is not the case that one ought not to be confident that *p*: one could not be justly reproached for being confident that *p*." The colon suggests that Ginet takes himself to be trivially rephrasing "ought not to be confident" with "would not be justly reproached for being confident". But this is not a trivial rephrasing!

use reliable processes, but he is intuitively not at all blameworthy in thinking that he is using reliable processes, and to form beliefs in the light of this belief. Classic process reliabilism entails that your internal twin's beliefs are not doxastically justified. Stewart Cohen thought this was a counterintuitive consequence, but he explicitly equated *rationality* with *justifiedness*. This is an extremely controversial equation—*at least* as controversial as the equation of justifiedness and responsibility. The natural move for the process reliabilist to make is to embrace a separability thesis. If he endorsed the weak separability thesis, he could claim that while the *belief* is *unjustified,* the *subject* is *not irresponsible*. If he endorsed the strong separability thesis, he could add that the agent is in fact praiseworthy. The process reliabilist can then note that this is precisely analogous to a plausible view in ethics. The resulting view is *prima facie* extensionally attractive.

4.1. *Hypological properties may not supervene on the current time-slice*

One might now get excited about the possibility of a compromise view: internalism is right about the hypological, and externalism is right about the deontic. This is too fast. To see why, we need to draw a distinction between *original* and *derivative* criticizability and appraisability.

The distinction is brought out in thinking about culpable ignorance in ethics. Suppose that you wittingly make yourself ignorant about the harmful effects of some act precisely because you know that you couldn't bring yourself to perform the act in full awareness of these effects in spite of wanting the act to be done *because* it has these effects. In later performing the act, you might believe—and indeed have good evidence for believing—that you are doing something perfectly fine. Are you blameless for your act? It is intuitive to think that you aren't.[77] Why not? Plausibly, because the reason why you are now in this cognitive position *vis-a-vis* your act is that this fits into an earlier plan you had for which you were clearly blameworthy. For this kind of reason, we cannot assume that if two agents are epistemic and motivational duplicates at *t*, they will also be hypological duplicates: we need to know about how they came to be as they are.

The same point applies in epistemology, and shows that the suggested compromise view fails. Suppose S realizes that if he doesn't cause himself destroy and forget the evidence for believing that his deceased wife didn't love him, he will end up believing it and become depressed. Everyone else will be nice to him and tell him that his wife did love him. So if he does cause himself to forget this evidence he should be able to believe accordingly. Suppose S succeeds in destroying and forgetting the evidence. Believing that his wife loved him is the correct response to his current evidence. Is he free from epistemic criticism in holding this belief? Plausibly, he isn't. This is a good example of derivative criticizability in the epistemic domain. Just as criticizability doesn't supervene on the current time-slice in ethics, so it doesn't in epistemology.

4.2. *How the separability theses could be false if epistemic Kantianism were true*

There is a deep connection between the separability theses and different views about our fundamental normative relation to truth as epistemic subjects. Strong epistemic Kantians who think that our fundamental normative relation to truth is the non-teleological one of *respect* ought to reject the separability theses. Strong teleologists ought to accept them. This is like a claim I'll eventually make about the deontic status of epistemic rationality: whether epistemic rationality is robustly deontic directly co-varies with whether one holds non-teleological or teleological views. Indeed, the two claims are more deeply related, since I think having respect for truth *partly consists in* satisfying requirements of epistemic rationality.

To bring out how epistemic Kantianism could imply the falsity of the separability theses, let's

---

[77]But see Smith (1983) for a classic discussion of reasons for doubt about the intuition.

consider first what respect *is*. A good idea comes from Darwall, who distinguishes what I want—which he calls *recognition respect*—from a different kind of respect he calls *appraisal respect*:

> There is a kind of respect which can have any of a number of different sorts of things as its object and which consists, most generally, in a disposition to weigh appropriately in one's deliberations some feature of the thing in question and to act accordingly. The law, someone's feelings, and social institutions with their positions and roles are examples of things which can be the object of this sort of respect. Since this kind of respect consists in giving appropriate consideration or recognition to some feature of its object in deliberating about what to do, I shall call it *recognition respect*.
>
> There is another attitude which differs importantly from recognition respect but which we likewise refer to by the term 'respect'. Unlike recognition respect, its exclusive objects are persons or features which are held to manifest their excellence as persons or as engaged in some specific pursuit. For example, one may have respect for someone's integrity, for someone's good qualities on the whole, or for someone as a musician. Such respect, then, consists in an attitude of positive appraisal of that person either as a person or as engaged in some particular pursuit. […] Because this sort of respect consists in a positive appraisal of a person, or his qualities, I shall call it *appraisal respect*.[78]

What is crucial to recognize about recognition respect is that it imposes a constraint on the subject's deliberation about what to do *in the light of her perspective on the facts*. To see what the constraint might involve, it is useful to consider an example from the practical case:

> *Ingenious Marital Reconnaissance.* Jane worries that her husband Dale would cheat on her if he had the opportunity. She decides to test this hypothesis in a clever way. Using her extraordinary costuming skills, she manages to dress up to look like a different woman on whom she suspects Dale would have an instant crush. Disguised, she starts regularly showing up outside his workplace to flirt with him. Dale believes on this misleading but compelling evidence that he is interacting with a different woman. He flirts with her in turn. Indeed, he makes plans to have a romantic evening with her when he next thinks that his wife is out of town, and not to tell his wife a word about it. That is what he thinks he is doing next as he goes through with his plans. He knows that his wife would be outraged if she found out. But he doesn't care, and shows his romantic interest to a great time.

Dale plausibly lacks the kind of respect that he ought to have for his wife. But why? He does not actually cheat on her—indeed, he does nice things for her, since she *is* the woman in disguise. His failure of respect cannot be located in any of his acts, understood as external occurrences involving other people. It is rather located in the relationship between his *motives* and *how things appear to him*—between various elements of his *perspective* on what he is doing. It is because he wittingly has strong evidence that he is doing something that violates the terms of his relationship with his wife, and because these epistemic facts have no constraining influence on his practical reasoning, that he fails to have sufficient respect for his wife.

This case shows how there could be a collapse of deontic facts into hypological facts. In the practical sphere, one usually draws the deontic-hypological distinction by claiming that while someone's evidence, beliefs, and motives may affect what she would be criticizable or creditworthy for doing, only her effects on the world affect whether she did what she *really* ought or ought not to do. If you have strong but misleading evidence that giving a certain liquid to a person would kill her when it will actually save her life, and she drinks it at your behest, there remains a clear sense in which *what* you do is right: you *save her life*, after all! But you are still plausibly open to criticism in acting, because you acted in a way that was wrong *by the lights of how things appeared to you*. So, the line goes, these evaluations are orthogonal: genuine rightness turns on the facts,

---

[78]Darwall (1977: 38–9)

whereas criticism of the agent turns on the agent's perspective on the facts.

But they cannot be orthogonal when what you ought to do is respect something: failing by the lights of how things appear to you *is* failing objectively in this case. Whether you count as respecting something or someone turns not just on your effects on the world, but on your perspective on them. Norms of respect cannot be blamelessly violated, or conformed with in ways that would not warrant at least some minimal positive appraisal.

If this is right, then even if it could be argued that certain apparent constraints in the epistemic domain are in the first instance standards for the criticism of the epistemic *subject*, this would not be enough to show that there are no significant corresponding genuinely deontic facts. If it it were independently plausible that the force of these standards is grounded in epistemic norms of respect, the distinction would simply collapse.

I think this points to what those with inclinations that are internalist about deontic properties like justifiedness must do. To avoid the obvious objection that they are conflating deontic and hypological categories, they can point out that there really wouldn't be a deep extensional gap between these categories if our fundamental epistemic obligation were the non-teleological one of respecting the truth. Then, they could argue for that claim.

5.   *The way forward: a preview*

That is precisely the kind of argument I will be giving in this book. Let me say a bit about why, and then sketch the purpose and strategy of the chapters to follow.

As I said in the Introduction, I started out as an internalist about justification, but became increasingly worried by thinking about some important analogies with ethics and the philosophy of practical reason. The first thing that worried me was that I was conflating two distinct properties: *rationality* and *justification*. The distinction between these is now taken for granted as deep fact in the philosophy of practical reason: what reasons there are—and hence what we are justified in doing—turns on the facts, but whether we are rational turns on how the facts appear. I think that, at least superficially, I *was* conflating these properties. Internalism *is* plausible about rationality, but almost certainly not the whole story about justification.

The second and much more important thing that hit me was that it is deeply unclear *how* epistemic rationality so understood could be seriously deontic. Many people in the philosophy of practical reason these days are skeptics about the idea that we really ought to do or even have any distinctive reasons to do what we are "rationally required" to do. There is just what we have most reason to do, and that comes apart from what it would be rational to do. Rationality is concerned not with the assessment of the *act* but rather with the assessment of the *agent*. If the same kind of conclusion extended to epistemology, then *even if* internalists were right about rationality, it would be a Pyrrhic victory: what it is and is not rational to believe may be *irrelevant* to justifiedness, the normative status that actually *matters* for thinking about what to believe.

For a while, I thought the conclusion did extend to epistemology, and so came to doubt that internalist factors could have *any* deontic impact in their own right. Surely, I thought, what matters in epistemology is true belief (and not having false beliefs). But rationality, if it is indeed internalist, is not a truth-conducive property. Indeed, it can be systematically truth-*obstructive*. That casts doubt on whether it has any epistemic value. After all, it clearly isn't *intrinsically* epistemically valuable. So there is room for doubt about whether we ought epistemically to care about it, and room for thinking that what we ought epistemically to care about is instead the property to which paradigm externalists about justifiedness point: reliability. Moreover, I doubted that it will help to

go pluralist and claim that knowledge is also intrinsically epistemically valuable. For one thing, it's unclear whether knowledge requires rationality. I think it doesn't. For another thing, I think complying with requirements of epistemic rationality could systematically diminish what we know. So the same worry arises even if we accepted a more expansive axiology—something I am, for independent reasons, extremely doubtful about doing.

What I realized is that this line of worry presupposes both weak and strong epistemic teleology. What I also realized is that if a non-teleological view about both epistemic value and the fundamental epistemic norm were taken, there would be a straightforward argument for thinking that epistemic rationality could be robustly deontic: this is the argument to which I began to gesture in the last section. My aim, then, shall be to try to vindicate the thought that internalist constraints are robustly deontic. It shall not be to argue for internalism about justification *per se*. Extensional considerations, together with the best view about respect, lead me to think that it cannot be the whole story. The point instead is to argue that what the internalist could be right about—epistemic rationality—always will have an impact on whether we are or are not justified in believing something. The strategy will be to settle the more fundamental debate between teleology and non-teleology, to argue for epistemic Kantianism, and to show how this vindicates the deontic robustness of requirements of epistemic rationality.

But first I need to justify some of the steps that led me to the initial worries. In Chapter 2, I'll defend the thought that epistemic rationality and justifiedness should be viewed as at least importantly conceptually distinct—so that it is an *open question* whether epistemic rationality suffices for justifiedness or *vice versa*—and also that a kind of internalism is right about epistemic rationality. In Chapter 3, I'll sketch some arguments for skepticism about the robustly deontic status of epistemic rationality, noting at the end of the chapter that these arguments stand or fall with strong epistemic teleology. In Chapter 4, I'll argue against both weak and strong epistemic teleology. Then, in Chapters 5 and 6, I'll defend epistemic Kantianism and explain how it vindicates the robustly deontic status of epistemic rationality, and hence of internalist constraints. This position has broader implications for some specific contemporary debates in epistemology— about internalism vs. externalism, the normative status of higher-order evidence, and how defeat works, *inter alia*. I'll discuss these in Chapter 8, after explaining the broader implications of my epistemic Kantianism for epistemic value theory in Chapter 7.

II

RATIONALITY

We can have been tempted […] only if we were blind to the richness of the stock of concepts that we must make use of in normative thinking. The idea that the concepts […] come to pretty much the same—a smooth, warm, conceptual pudding—is just a mistake.

—Judy Thomson[79]

1.   *Introductory remarks*

I turn to discuss rationality. At first I'll keep the discussion general, and point to core facts about rationality that hold in a domain-neutral way. Then I'll discuss views about the relations between *practical* rationality and normative reasons. The view I favor will have a precise analogue in epistemology, to which I'll then turn. That is one main reason for discussing practical rationality. I'll also need my views on it in place to make key structural analogies clear in the next chapter.

As we'll see, while I endorse the increasingly popular idea in the practical literature of separating rationality from responsiveness to genuine reasons, I think many in this literature are mistaken about what rationality is and how it differs from responsiveness to genuine reasons. To have a convincing case that rationality in the practical and epistemic spheres has a common core, I must correct these mistakes. This is something I've independently done, though I now see the details a bit differently.[80] If I went with the majority of philosophers of practical reason and proposed a *precisely symmetrical* view to theirs in epistemology, the result would be either false or have to be about a less important concept that they have often conflated with rationality—what I'll call *intelligibility*, which is closely tied up with sense-making explanations of intentional behavior.

A virtue of my overall view is its unifying character. This type of ambition is not unusual in the new literature on rationality, though epistemic rationality has not yet been given the level of special attention it deserves. While the best of this literature is written by ethicists and philosophers of practical reason, many of them write with domain-neutral ambitions. That is one reason why epistemologists ought to read this literature. The quality of the discussion of the general features of rationality exceeds that of discussions by epistemologists. Epistemologists have not often been sensitive enough to the right distinctions, and have often failed to discuss rationality in its own right. They can learn from this literature.[81]

---

[79]   Thomson (2008: 94).

[80]See Sylvan (ms₁).

[81]There are epistemologists working on intersection questions who may claim to be interested in practical rationality—e.g., Stanley and Hawthorne (2008). But they are best understood as discussing some intermediate property. There is no plausibility in the idea that if one fails to know that P *only* because one is Gettiered, one is irrational, or less than fully rational, in acting for the reason that P. Putative knowledge norms on action and practical reasoning may be genuine norms, but they are not norms of rationality. The typical reply by Stanley and Hawthorne to the concern about ignorance due to mere Gettierization brings this out. They say the agent *excusably violates* their norm. Excusable violation is often a fine idea, but not for rationality, as I'll insist when I get to what I call "act-agent blur" in §1.2. Stanley and Hawthorne have in mind something more like what Scanlon calls *reasonableness*, a standing he took to be in between rationality and responsiveness to reasons.

1.1.   *General features: narrowness and perspectival character*

The first thing to emphasize about rationality in the most familiar sense is what Scanlon calls its **narrowness**—though it is not, I think, *as narrow* as he thinks it paradigmatically is.[82] To see this, recall again the case from Williams I discussed in the Introduction. Bernie is in his favorite bar. He ordered some gin and tonic. Alas, the bartender is in a quietly sociopathic mood tonight, and puts petrol in the glass instead. This is visually indistinguishable from Bernie's favorite drink. The fact that the glass contains petrol is a conclusive reason for Bernie not to plan to take a sip. Still, Bernie is surely not *irrational* for planning to take a sip. Indeed, Bernie seems perfectly rational for planning to do so. The two observations go hand in hand. Rationality in the most familiar sense is the positive counterpart of *irrationality,* and criticisms of irrationality are clearly narrow. Cases like this—which can be indefinitely multiplied and varied—suggest that forming rational intentions or acting rationally demands less than responding to the genuine normative reasons bearing on intentions and acts in virtue of one's circumstances.

While one can stipulatively use 'rational' and 'irrational' differently, this is the most natural way to understand the phenomenon such terms are used to identify. The recent literature on rationality by ethicists and philosophers of practical reason I mentioned earlier is explicitly about this natural, narrow sense. It is what interests me. To disregard the narrowness of rationality is, I'll assume, to change the subject, though one might have a different picture about *wherein* the narrowness consists. Typically people in this literature agree that there are less narrow ideas—e.g., responding to genuine reasons—but insist on separation, usually well in line with ordinary thinking as it is revealed in cases like Williams's "petrol and tonic" case. Exactly what the relevantly parallel distinctions are in epistemology is something I'll address later in this chapter.

It is worth stressing that rationality is not *simply* narrower than responsiveness to *sufficient reasons that exist*. It is also narrower than responsiveness to *possessed sufficient reasons*. At the very least, it should be *controversial* whether rationality in A-ing and correct responsiveness to possessed sufficient reasons in A-ing coincide. I'll point to two connected reasons for thinking so, and respond to some (bad) objections that will naturally arise along the way.

Consider:

> (**Boring**) Possessed sufficient reasons = sufficient reasons that are, by the way, possessed.

(Boring) looks to me like a platitude about what it is to possess sufficient reasons. What else could it be? God—or Mark Schroeder—only knows.[83] At any rate, (Boring) is plausibly a constraint on any *Factoring Account* on which *possessing reasons* is a matter of *there being genuine reasons to which you bear a possession relation*. The claim "You have sufficient reasons to A, but there are not sufficient reasons to A" cannot be read coherently, and it is a virtue of the Factoring Account that it can explain this

---

[82]"Narrow" here is not being used in the sense familiar from internalism/externalism debates in the philosophy of mind.   There is nothing *incoherent* in a view that makes rationality in the ordinary sense turn on factors other than a subject's nonfactive mental states, though I'd encourage us to reject such a view. It is a just a term taken from Scanlon whose sense is fixed by examples and contrasts with what we take to be less narrow—viz., responsiveness to (perhaps possessed) normative reasons, and maybe some further ideas. Scanlon does deliberately fixate on an *extremely* narrow idea—one self-consciously much narrower than the one used by others who also make a reasons/rationality distinction (e.g., Parfit). But if we're simply interested in the contrast between responsiveness to genuine reasons and rationality in the familiar sense, I think this fixation is unhelpful and far too narrow. I will elaborate on this at length in §2.

[83]Schroeder (2008) denies that having a reason is a matter of bearing a possession relation to a reason. Revisit footnote 26 for more discussion. Again, I think he is wrong about this, and that the Factoring Account is true. See Lord (2010) for one excellent critical discussion of Schroeder's view.

so transparently. Indeed, to bring out how natural it is to cleave rationality from responsiveness to possessed sufficient reasons, suppose it appears that there are genuine sufficient reasons for A-ing, and that this appearance would be sustained after indefinite reflection on the facts to which you have access. Surely this appearance could still be misleading, so that there aren't really sufficient reasons for A-ing. It is not as if there is no appearance/reality distinction for possessed sufficient reasons! Supposing otherwise makes a mockery of normative inquiry. If there aren't *really* sufficient reasons for A-ing, there *ipso facto* can't really be possessed sufficient reasons for A-ing, assuming (Boring). Still, it is very intuitive that it is rational for you to A here. So, rationality doesn't imply having sufficient reasons.

Here is another, stronger reason for denying that rationality just is correct responsiveness to possessed reasons. Assuming the Factoring Account, you cannot possess a consideration as a normative reason unless the consideration possessed is actually a real normative reason. But the following case is conceivable. Sometimes there are *enabling conditions* for reasons—conditions that must be satisfied for considerations to count as real normative reasons. Consider a case in which—by the lights of the appearances—all the enabling conditions for R are satisfied, where R would be a strong reason if they were satisfied. The appearances mislead: the enabling conditions are not satisfied. It seems that it would be irrational for you to *fail to respond to R exactly as if it were a real normative reason*, and that it would be perfectly rational for you *to respond to R exactly as if it were a real normative reason*. Still, R is not actually a reason at all, because the appearances mislead: R's enabling conditions aren't satisfied. *A fortiori*—assuming the Factoring Account—R is not a reason you have. Suppose you do respond to R in the way you ought if the enabling conditions were satisfied. You are perfectly rational in doing so. That can't be because R is a reason you have. R isn't a reason at all. R would have to be a reason to be one you have.

One could retreat to the claim that there are *other* reasons you have that make it rational to respond as if R is a real reason here. Perhaps the *very fact* that the total appearances suggest that R's enabling conditions are satisfied is itself a reason to respond as if R were a real reason. But now the problem is one of *degree*. Call the different reason appealed to by this strategy "the Appearance Fact". Even if the Appearance Fact is a reason to respond as if R is a real reason, it is not *as good* a reason to respond *exactly as if* R is a real reason as R itself would be. How could a misleading appearance of a reason be *as strong a reason* for responding as the reason itself? I say it couldn't be. Otherwise there is no appearance/reality distinction for normative force. So, on this view, if the Appearance Fact is a less weighty reason than R itself, then responding to R *exactly as one would respond if R were a reason* will *not* be *as rational* as it would be if the appearances *weren't* misleading. But that is wrong. Whether or not the appearances mislead makes no difference *per se* to one's degree of rationality. So the retreat does not really help.

For these reasons, I think it is a mistake to presuppose as a point of analytic bookkeeping that rationality and correctly responding to possessed reasons coincide. These things can come apart in deep ways. Rationality will be narrower than responding to possessed reasons unless one defends a narrower view of reasons than the one I invoked, gives up the Factoring Account, understands possession in a surprising way, or understands the factoring in a surprising way.[84]

---

[84]One idea might be that "sufficient" is to be understood in two different ways depending on whether it is embedded in the "have" context. Consider a specific Factoring Account to see how this might go: Errol Lord's view on which *having a reason* is *being in a position to know of a fact that is a reason that it obtains*. The *being in a position to know* relation is like a spotlight. One might think that one can collect together the reasons in the spotlight, and consider how these weigh against each other *considered just on their own*, *as though the spotlit world were the whole world*. Suppose one is only in a position to know about two reasons bearing on whether to A—one a reason R of a certain strength for A-ing, one a reason R* of slightly greater strength against A-ing. If we focus just on these, and pretend that they are all the reasons there are, one might say: "You lack a sufficient reason for A-ing, since R is outweighed by R*." The idea, then, is that when embedded, what is intended by

So much for mere narrowness. A more informative though less neutral way to summarize the observations about cases like Bernie's is to say that whether someone is rational or irrational in a given case turns not on the totality of facts about her circumstances that give normative reasons, but rather on the *perspective* available to her on what these facts are. We can call this the **perspectival character of rationality**. Exactly how we ought to understand this is a difficult question. The bare idea, however, is an extremely plausible core thought about rationality.

As I'll discuss more in the next subsection, an orthodox line in the *practical* case is that 'perspectival' should be understood as 'belief-relative'. As Parfit says: "While reasons are provided by the facts, the rationality of our desires and acts depends instead on our beliefs."[85] Scanlon says something similar but stronger: "Irrationality in the clearest sense occurs when a person's attitudes *fail to conform to his or her own judgments*: when, for example, a person continues to believe something […] even though she judges there to be good reason for rejecting it […]. These are clear cases of irrationality because the thought or action they involve is in an obvious sense 'contrary to (the person's own) reason': there is a direct clash between the judgments a person makes and the judgments required by the attitudes she has or holds."[86] This kind of view may have its origins in Davidson, who said that "irrationality is a matter of inner inconsistency", where the inconsistency is between the agent's attitudes and her acceptance of certain principles constitutively "shared by all creatures that have propositional attitudes or act intentionally".[87]

Crucially, this is not the only way to understand the perspectival character of rationality. A few suggest that rationality in the narrow sense is *evidence-relative*: the perspective is the perspective of the person's total evidence about the reasons she has.[88] While I once argued for a version of this view, the view that now seems to me to be best is that rationality is *appearance-relative*, where 'appearance' is understood in a *nondoxastic* sense. The perspective is the one created by these nondoxastic appearances on what reason-giving facts exist. More on this in a few subsections.

1.2. *General features*: *local person-orientation and act/agent blur*

The second general thing I would note about rationality is not always emphasized enough. Talk about rationality applies in the first instance to *persons* in specific cases *for* acts or attitudes. Talk acts or *attitudes* being rational or irrational is *derivative*. Parfit puts this nicely: "When we are trying to decide what we or others ought to do, what matters are the reason-giving facts […]. When we ask whether someone has acted rationally, we have a different aim. We are asking whether *this*

"sufficient" is a measure of how the reasons *within the spotlight* weigh up against each other alone. Clearly, it could be that there is a further unlit *intensifier* for R—a consideration outside the spotlight that is not itself a reason for A-ing that, when it obtains, makes it the case that R has extra weight. (See Dancy (2004a: 42) for discussion of intensifiers if you are unfamiliar with them.) If one added the intensifier to the previous collection, it would make R tip the balance in favor of A-ing. Suppose this intensifier is the only other relevant fact that could be added to the balance of considerations bearing on whether to A in the entire world, and *isn't* added to the spotlit collection. Considering things from a God's-eye point of view, R is a sufficient reason for A-ing. Here it might be true that R is a reason you have, and a sufficient reason, without your having a sufficient reason. When "sufficient" gets embedded within "have", the calculation is different than the one that results when it occurs outside.

    This "double-thinking" view about sufficiency may help. But I wonder whether it makes sense. If "sufficient" is shifty, it should be possible to hear the following as a coherent sequence, where 'reason' is understood in the same sense in all of them, as on any Factoring Account: "R is a sufficient reason for A-ing. You have R. But you do not have a sufficient reason for A-ing." I cannot hear this as coherent.

[85] Parfit (2001: 17). As I'll stress in a later footnote, Parfit may not really accept a purely belief-relative view.
[86] Scanlon (1998: 25).
[87] Davidson (2004: 194–5).
[88] Lord (Ms$_1$) and Ross (forthcoming) both hold views of this kind.

person deserves the kind of criticism that we express with words like 'foolish', 'stupid', and 'crazy'."[89] Plausibly, then, the conceptually basic types of rationality claims have the forms:

(a)   S is rational in A-ing   and   S is irrational in B-ing
(b)   S would be rational in A-ing   and   S would be irrational in B-ing

Claims like (a–b) are structurally akin to paradigmatically person-oriented hypological claims like "S would be praiseworthy in A-ing" and "S is blameworthy in B-ing". We *can* also talk about

(c)   whether rationality requires A-ing,   and
(d)   whether rationality allows B-ing.

But this is, I believe, derivative talk. Consider an analogy with a kind of evaluation that—as Parfit rightly hints—evaluations of rationality resemble: specific-case evaluations of whether someone is behaving *intelligently*. Notice that we can say things like "Intelligence in B-ing requires A-ing". Surely what this means is not that intelligence is a *source* or *code* of requirements, akin to the law or etiquette or an instruction manual, which calls for A-ing when one is trying to B. What this means is that one is assessable as B-ing fully intelligently *only if* one also does A.

Something similar plausibly holds for (c–d). "Rationality requires S to A" is true in a choice situation C iff S would be irrational for failing to A in C, and "rationality allows S to A" is true in a choice situation C iff S would not be irrational in A-ing in C. The fundamental sense is not what Broome calls the "*source*" or "*code*" sense, but rather the *property sense*—though the property is a *local* one relativized to the person *in doing* a specific act or attitude or configuration of attitudes, and rather than the property of *being rational*, all things considered.[90] It is not *ad hoc* to think there is such a distinction. The same local/global distinction shows up for other hypological assessments. We can talk about whether one is blameworthy *in A-ing*, but also about whether one is blameworthy *period*, which may turn on whether the person has defective dispositions.

This vindicates Kolodny's observation that rational requirements are "local", and focused on specific conflicts in one's mental life.[91] We say that rationality prohibits a conflict for one when,

---

[89]   Parfit (2011: 36); italics mine.

[90] Broome (2007a) explores the logic of the code sense, but admits that the property sense exists. In Broome (2007b: 362), he says that "the code sense is more natural than the property sense", but doesn't explain why. Of course, since he and Kolodny are arguing about collections of specific requirements, it is nice to think of there being a "code of rationality" that lists them. But if we ask *where* these requirements come from, it is natural to advert to the property sense. This is clearly *not* what we do with legal requirements. With rationality, local evaluations of people come most easily. Our intuitions are fundamentally tracking whether someone has a local property, rather than whether she complies with some independent rule. So, together with the analogy with "requirements of intelligence", I see sufficient reasons to prefer my view.

[91] Here was Kolodny's original observation:

> Our ordinary judgments about rationality […] are local. They are focused on specific conflicts among one's attitudes. We might judge, for example, that a person is being weak-willed in believing that he has conclusive reason to X, but not intending to X. And we might judge, at the same time, that he is giving into wishful thinking in believing that he has conclusive reason to X, having himself decided that the evidence for that belief is flimsy. Rational requirements, accordingly, ought to be local. In each case in which one is under a rational requirement, what it ought to require of one is to avoid or resolve some more specific conflict among one's attitudes—as opposed, say, to satisfy some global constraint on all of one's attitudes. […] Various applications of rational requirements may call for one to form some attitudes, to retain others, and to revise still others all at the same time." (2005: 515–6)

> Not all of this is obvious: what is going on globally in someone's mental life might plausibly influence whether there is sufficient reason for regarding her as irrational *in instantiating a local configuration of states*. But setting that aside, the point for us is that taking the property sense as basic does not force us to reject Kolodny's idea.

if all we knew is that one suffers from this conflict in some case C, we would have sufficient reasons for criticizing one as irrational in C *for this conflict*. This puts the agent-oriented criticism first in a localized way, without thinning it into a broader comment on whether one possesses or exhibits the faculty of rationality, or has a good or bad rational dispositions.

We hereby see how evaluations of rationality are in the first instance person-oriented and hypological. But it is hard to think that this is *all* there is to our thinking about rationality. Rationality also seems to have a deontic side, as Kolodny observed:

> It is not always true that when we claim that someone is irrational, we are only appraising him. Sometimes we seem to be saying something […] deontic, or response guiding: something in the register of advice, rather than [mere] assessment. An atheist might say to a racist believer: 'Look: I think you're nuts to believe in God, let alone that He created anyone. But given that you believe that God created all people equal, and given that you agree that people whose skin is a different color from yours are people, you *ought* to believe that He created them equal too. It would be irrational of you not to.' Claims like this do not seem to be grading the addressee […].
>   Consider […] this normative dimension from the inside. Often when we ourselves are subject to rational requirements, we feel that we ought to respond as they require, or at least we can be brought to feel this when the right things are called to our attention. […] [O]f the response that rationality requires of us, we think that we ought to give it […]. [T]his thought is something distinctively normative, rather than evaluative […]. In sum, there is a kind of 'ought' that comes naturally to us when subjective rationality is at stake, and this 'ought' seems, like the 'ought' of having conclusive reasons, to express something normative rather than evaluative.[92]

So even if the local property sense of 'rationality requires' is the fundamental sense, we would be missing something if we left it there. These criticisms are naturally accompanied by a feeling and a suggestion of real deontic pressure. Failing to respond to them seems to involve running afoul of a genuine demand. An adequate theory of rationality must either explain this or explain it away. This desideratum is not easily satisfied, which leads to the **problem of deontic significance** I'll discuss in Chapter 3, and generalize for epistemic rationality.

Another general feature of rationality worth noting accompanies the last two points. Recall that, at the end of the last chapter, I noted how many people accept **separability theses** about the deontic and the hypological for some domains: one can violate a moral requirement blamelessly, and conform with it and deserve no praise for this. The fact that the local property sense of 'rational' is the fundamental one, and that particular requirements of rationality are introduced to ground local person-oriented appraisals, suggests that this will not be true of rationality. While rationality does have an apparent deontic side, it seems inextricably bound up with the hypological side. This doesn't mean that there was never a deontic side to begin with. It just indicates that separability theses fail for rationality. Unless we switch to talk of responsiveness to reasons, there is little sense in the idea that one can rationally excusably violate a requirement of rationality, or comply with a requirement of rationality and fail to be rationally creditworthy for it. There is thus what I call "**act-agent blur**" for rationality. The deontic and the hypological seem to be blurred in a way not true of other domains (e.g., moral evaluation).

As I'll explain further later, I think this apparent fact about rationality explains why internalists

---

We can engage in isolation: assume everything else is fine, and consider whether having a configuration of mental states *would give reasons to rationally criticize in the particular case*. So the disagreement over locality can just be about the nature of the relevant property we need to ground the requirements in the property sense. It is thus odd when Broome (2007b: 362) claims that Kolodny and he agree that the code sense is fundamental *simply* by noting that Kolodny endorses locality. Locality underdetermines which of the two senses is fundamental.

[92] Kolodny (2005: 554–5).

often assume that there is a deep link between epistemic justification and responsibility—an assumption that otherwise looks confused when we notice the plausibility of separability theses for other domains. Just recall BonJour: "My contention here is that the idea of avoiding such irresponsibility, of being epistemically responsible in one's believings, is the core of the notion of epistemic justification."[93] Notably, internalists like BonJour *also* tend to presuppose that epistemic justification and rationality are identical. If that were so, the fact that separability theses plausibly *do* fail for rationality would explain why they are so tempted to go on to tie their intuitions about epistemic justification to intuitions about doxastic responsibility.

1.3.   *General features*: ex ante *and* ex post *evaluations*

A final general fact about our thinking about rationality is worth noting. In thinking about both practical and epistemic rationality, we can distinguish **ex ante** and **ex post** evaluations.[94] One cannot think clearly about rationality without being careful to mark which one has in mind.

What is the distinction? *Before* someone acts or forms some attitude, we can discover truths about whether this person *would be rational* in doing so. Suppose Jones is walking, and has a visual experience exactly like the one he would have if there were a tree ten feet away. Were there such a tree, he'd run into it if he didn't move left. Here we can say that Jones would be rational to move left irrespective of whether his visual experience is accurate. This is an *ex ante* evaluation.

After Jones acts, we can consider more. We can consider how the factors that made it *ex ante* rational for him to move *hook up* with his movement. Suppose Jones moves left, but that his movement is not at all related to the apparent fact that he would otherwise run into a tree. He moves left because of some superstition of his. Were it not for this superstition, he would keep moving in the direction where there visually seems to him to be a tree. It is hard to regard Jones as fully rational *in moving left* here. But why? Surely it is rational *ex ante* for him to move left, given the visual appearances. While he is rationally criticizable for the superstitious beliefs he has, this could be a separate matter from his irrationality *in moving*. The best explanation of why Jones is not fully rational *in moving* is that the facts that made it true that it *would be* rational in the *ex ante* sense for him to move *failed to explain* his movement. He failed to move *on account of the factors that made it rational* ex ante. This is why, though it is *ex ante* rational for him to move, there is a different sense in which he was not rational in moving.

This is related to my distinction between *having a good reason for A-ing* and *A-ing for it in the strong sense* from the last chapter. There is also an analogy with the distinction between propositional and doxastic justification. Jones is like someone who possesses good reasons for belief, but simply fails to believe for these reasons. I'll return to this in §2.5, since the distinctions will turn out to be more closely related, even though they are not exactly similar.

2.   *Orthodox apparent reasons views of perspectival character*

How should we explain the perspectival character of rationality that seems to be illustrated by cases like the original Bernie case? A tempting generic thing to say is that how things *appeared* to

---

[93]   BonJour (1985: 8).

[94] I take this terminology from Goldman (1979), who used it to mark the contrast between propositional and doxastic justification. As Wedgwood (ms) notes, this terminology is really better, since there is a much more general distinction to be made of which the propositional/doxastic distinction is just one example. We might miss out on the important structural similarities across domains if we insisted on using the propositional/doxastic language in epistemology. This is, I believe, true of Turri (2010), who couldn't have thought doxastic justification can't be analyzed in terms of propositional justification and purely nonnormative properties and relations if he had considered other cases where the same distinction arises.

Bernie explained why he was *ex ante* rational in intending to take a sip. More specifically: the fact that the reasons that appeared to exist did not count against taking a sip explains why it is *ex ante* rational for Bernie to sip. Call this generic thought the **Apparent Reasons Theory**. The orthodox version of this theory can be called the **Pure Doxastic View**. It takes things one step further: in saying that how things *appear* fixes the facts of rationality, we are using the *doxastic* sense of 'appear'.[95] Notably, Pure Doxastic Views don't claim that the only factors that influence practical rationality are beliefs. They claim that the only factors that influence rationality *other than paradigmatic motivational states like desire and intention* are beliefs. There are, of course, other requirements of coherence involving paradigmatic motivational states. There is the *Subjective Instrumental Principle*: rationality requires that if S intends to A, and it appears that B-ing is necessary for A-ing, S intends to B. There is also *Intention Consistency*: it is a requirement of rationality that if one intends to A, one's A-ing is consistent with the fulfillment of one's other intentions, relative to the appearances. Pure Doxastic Views are about rationality and irrationality in intention or action, *setting aside* conflicts with paradigmatically motivational states.

There are two general versions of this view. One version inspired by Scanlon (1998) is the

> **Weak Doxastic View**: Setting aside conflicts with other paradigmatic motivational attitudes, S would be *ex ante* rational in intending to φ (or in φ-ing) iff S believes that there are (at least) sufficient reasons for φ-ing, and S would be at least somewhat *ex ante* irrational in intending to φ (or in φ-ing) iff S believes that there are sufficient reasons against φ-ing.[96]

Scanlon supported something like this view by noting that people are most clearly irrational when they act against what they acknowledge to be reasons, and seem rational in proportion to the extent to which they conform to what they acknowledge to be reasons. What is distinctive about this view is that the beliefs that determine what someone is rational or irrational in doing are *explicitly normative*. Other friends of the Pure Doxastic View reject this claim, and hold the

> **Strong Doxastic View**: Ignoring conflicts with other paradigmatic motivational attitudes, S would be *ex ante* rational in intending to φ (or in φ-ing) iff S has beliefs whose truth would give her sufficient reasons for φ-ing, and it would be at least somewhat *ex ante* irrational for S to intend to φ (or to φ) iff there would be sufficient reasons not to φ were S's beliefs true.[97]

---

[95] The thought that there are many different senses of 'appear' and the related 'look' is common among epistemologists and philosophers of perception. It goes back at least to Chisholm (1957), and was recently nicely discussed by Alston (2002b) and Lyons (2005).

[96] *Nota bene*: rationality and irrationality aren't exhaustive categories. This account does not assume otherwise. As Hursthouse (1991) says, people may sometimes be *arational* in acting. If someone is incapable of telling whether she has or lacks sufficient reasons, it is unclear that we ought to call her rational. We also shouldn't call her irrational. Exactly *which* beings (and *when* beings) count as *arational* may depend on what it takes to have the concept of a reason. I think it doesn't take much. Any creature capable of *responding* to reasons rather than *merely reacting* to them has it. And any creature who is disposed to treat to certain considerations as reasons counts as having some normative beliefs about the bearing of those considerations on action. It may not be able to *articulate* its normative beliefs, or use the *word* 'reason'. But concept possession does not generally require such articulability, or that one to know how to use the related words. More on this in §2.7.

[97] I should note that this differs from Parfit's own view, which is subtler. The subtlety is brought out in part by the fact that he uses the word '*clear*' to describe the belief-relatively apparent reasons that matter most for rationality. 'Clear', like 'appears', can be understood in different ways. It can point to what would most strongly seem true *in the light of the agent's total set of beliefs*. It could also point to what *ought epistemically* to seem true to the agent, given the agent's total mental state. And it could also be understood as a claim about what the person's nondoxastic representational states—her experiences and intuitions—strongly suggest to be true. Obviously, if 'clear' is understood in the second or the third way, Parfit's view is not a version of the Pure Doxastic View. And if it is understood in the first way, his view collapses into the very Weak Doxastic View he rejects. This reveals, I think, that he doesn't actually believe any Pure Doxastic View. He has confirmed

One might think that the fact that this view does not make rationality turn on explicitly normative beliefs is a virtue. Something Parfit says can be used to support this thought:

> People are most *clearly* irrational, Scanlon claims, when they fail to respond to what they themselves acknowledge to be reasons. This claim is in one way true, since such people are less than fully rational even according to their own beliefs. If these people were charged with not being fully rational, they would plead guilty. But that does not justify the claim that only such people should be called irrational. On Scanlon's view, even if we often fail to respond to very clear and decisive reasons, we could avoid irrationality merely by having no beliefs, or false beliefs, about which facts give us reasons […]. We ought, I believe, to reject this claim.[98]

As Parfit notes, the distinctive feature of Scanlon's view makes practical irrationality implausibly easy to avoid. On the Strong Doxastic View, it is harder. On this view, as long as one acts out of line with the reasons that one *would* have if one's *nonnormative factual beliefs were true*, one would be irrational regardless of one's explicitly normative beliefs.

But I think this is only a superficial advantage. The Strong Doxastic View actually makes rationality *too hard* to achieve. Indeed, I'll argue that *any* non-arbitrary version of the Pure Doxastic View is necessarily either like the Strong or Weak Doxastic View in making rationality either too hard or too easy to achieve. So we must reject all Pure Doxastic Views.

2.1.   *A dilemma for purely belief-relative views*

Here is the simple dilemma against Pure Doxastic Views whose horns I'll extend in this section:

1.   On the Weak Doxastic View, practical rationality is implausibly easy to achieve.
2.   On the Strong Doxastic View, practical rationality is implausibly difficult to achieve.
3.   Every non-arbitrary Pure Doxastic View will have either the property that made rationality too easy on the Weak Doxastic View or the property that made rationality too hard on the Strong Doxastic View.
4.   So, no non-arbitrary Pure Doxastic View is plausible.

We have already seen why premise (1) is true in considering Parfit's objection to Scanlon. But it is worthwhile adding a few remarks here. The objection trades on the fact that straightforwardly false and crazy normative beliefs are easy to maintain if one is sufficiently dogmatic or thick-skulled in ignoring intuitively obvious normative truths. The Weak Doxastic View implies that acting on such beliefs may be perfectly rational. Practical rationality is not so easily achieved.

We are not changing the subject in advancing this objection. One could *stipulate* a more demanding notion of ideal objective rationality that coincides with responsiveness to truly normative reasons.[99] The objection does not make the boring claim that the ordinary narrow sense of 'rational' is too narrow and that it is more useful to use 'rational' in this stipulative way. Doing that would change the subject. The objection concedes that rationality in the ordinary sense is narrow. It just insists that it is not *as narrow* as the Weak Doxastic View would have us believe. This is not implausible. If minimal reflection would make it intuitively obvious that φ-ing is disfavored by sufficient reasons and one sustains the belief that φ-ing is favored by sufficient reasons only by pigheadedness or dogmatism, one wouldn't be rational in φ-ing.

It is worth noting here that I concede that there is an even narrower concept for which an analogue

---

this in conversation.
[98]   Parfit (2011: 122).
[99]   Scanlon approves of such a notion; see his (1998: 30−2).

of the Weak Doxastic View would be defensible. But this concept is not that of rationality. It is the concept of **intelligibility**. There is an evaluative difference between someone who behaves randomly and complies with no standards, not even crazy ones, and someone who subjects her behavior to standards, including crazy ones. The former person's behavior is *unintelligible.* We cannot *make sense of it* at all: no coherent explanation in terms of motivating reasons can be given. But we might be able to make sense of the acts of an irrational person by taking a look at her motivating reasons, even in acknowledging that these are so crazy as to make her clearly irrational in acting. Mere intelligibility does not suffice for rationality even in the ordinary narrow sense. Once this distinction is drawn, it is clear that we ought to reject the Weak Doxastic View, since it only plausibly applies to intelligibility. I will return to the topic of how rationality and intelligibility differ in the next two subsections.

The rationale for premise (2) proceeds from this simple observation: excusable normative ignorance can be extensive. This is particularly vivid when the normative claims we are considering are specific ones of the form "State of affairs S would, if it obtained, give sufficient reason to φ". We may fail to know many truths of this form, and fail even to be in a position to know them. Take any such truth. We might still truly and sensibly believe that the state of affairs S obtains. When we do, the following type of claim will also, *unbeknownst to us,* be true:

> **Claim Schema**: We have a belief B which would give us sufficient reason to φ if true.

The Strong Doxastic View entails that the truth of instances of the Claim Schema is relevant for assessing whether we φ rationally. Not so. If we are totally in the dark about whether there would be a reason for A-ing were B true, we needn't be less than fully rational in failing to φ.

There is a simple formula for generating these cases. Take any intelligent human agent in the middle of difficult moral reflection. There are probably lots of unsettled considerations bearing on whether she could rationally believe different moral theories that disagree strongly about cases. She is not in a position to *settle* on any theory. Take a potential act A of hers on which one moral theory T—say, act utilitarianism—yields the verdict *permissible* while another theory T*—say, rule utilitarianism—yields the verdict *impermissible,* and suppose that self-interested or partial reasons would be too weak to dampen the force of the reasons that would disfavor A-ing if T* were true. Suppose her degrees of rationality in believing T and T* are close to equal, but that she would be barely more rational in believing T. *Also* suppose that T* is the true theory. If so, then the fact that an act she is considering is an A-type act is a decisive reason against doing it. But it is implausible to say that she is practically irrational. Unless one held the remarkable view that intuitive plausibility can never distribute inconclusively over conflicting theories, the formula can be used to generate lots of cases that suffice to refute the Strong Doxastic View.

In thinking about this, it is crucial to remember that the defender of the Strong Doxastic View cannot appeal to effects of *mere* normative beliefs on rationality in acting on pain of making rationality too easy. Only beliefs about the *nonnormative facts that give reasons* are supposed to matter on this view. After all, a big reason for moving from the Weak to the Strong Doxastic View was precisely that one cannot intuitively achieve rationality simply by changing one's normative beliefs. To revise the view in the imagined way would collapse it back into the Weak Doxastic View. Moreover, the defender of the Strong Doxastic View also cannot appeal to facts about the clarity of reasons. If 'clear' were understood doxastically, we would be back with the problem for the Weak Doxastic View. If 'clear' were understood epistemically normatively or in terms of nondoxastic appearances, the view would cease to be a Pure Doxastic View.

So I find it hard to see how the Strong Doxastic View can be modified so as not to mistake sensible actions under excusable normative ignorance for shortcomings on the score of practical rationality.

This is especially obvious if we hold—as I think we ought—a buck-passing account that reduces facts about *value* to facts about natural properties giving reasons. One *might* have wanted to respond by saying that, in cases of excusable normative ignorance, we are unaware of some features would that play a crucial role in explaining why we would have the reasons we would have: namely, some purely evaluative features. If the natural facts are what provide the reasons, this response fails. We may know all the natural facts but be rationally unsure about reason-giving relations. Such relations are often unclear. (Ethics is hard!)

The distinctive feature of the Strong Doxastic View that seemed to give it an advantage over the Weak Doxastic View is actually its key flaw. Nothing *follows* about whether someone would be rational or irrational in φ-ing from facts about her beliefs about the nonnormative facts that give reasons. Something *follows* only if we add facts about how these facts *normatively look* from this person's perspective. Crucially, however, this is no argument for returning to the Weak Doxastic View. For as we'll see, 'look', 'appear', and so on, do not need to be understood doxastically.

2.2. *Underived enkratic principles? Yes: they are wide-scope*

Now, you might think that departing from Pure Doxastic Views will prevent us from explaining simple facts about rationality. Wasn't Scanlon right that someone is clearly irrational if she believes that she has decisive reasons to φ and fails to intend to φ, and if she believes that she has decisive reasons not to φ and intends to φ? Well, yes. But the denial of these claims does not need to be built into the rejection of Pure Doxastic Views.

There are other requirements of rationality besides the requirement to respond to apparent reasons in whatever nondoxastic sense of 'apparent' is best for capturing the perspectival character and narrowness of rationality. Some of these are **wide-scope coherence constraints**. Earlier I distinguished between *rationality* and *intelligibility*, and said that intelligibility does not suffice for rationality. The converse is not true: rationality *does* suffice for intelligibility.

As I see it, wide-scope constraints are the principal requirements of intelligibility. These requirements merely ban us from being in incoherent states. The requirement to respond to apparent reasons in some appropriate sense of 'apparent' is a separate narrow-scope requirement that is independent of these requirements. My full view is that practical rationality requires complying with this requirement, with some version of the Subjective Instrumental Principle, some version of the requirement of Intention Consistency, and with wide-scope constraints like:

> (**A – Enkrasia WS-1**)   For any S, rationality requires S not to simultaneously believe that there are decisive reasons to φ and fail to intend to φ.

> (**B – Enkrasia-WS 2**)   For any S, rationality requires S not to simultaneously believe that there are sufficient reasons not to φ and to intend to φ.

This is a perfectly consistent position.

Of course, some doubt whether (A) and (B) could be requirements. Kolodny, for instance, tries to argue that there are no requirements such as (A) and (B).[100] I think his main arguments for this claim fail, though I agree with him that some requirements of rationality are not wide scope. We do not have to take a uniform position about the form of requirements of rationality, and claim that *all* requirements are either wide-scope or narrow-scope. The semantics of requirement talk might pressure us to be *superficially* uniform. But the dispute here is not about semantics. It is

---

[100]   See Kolodny (2005), (2008a), (2008b) and (2008c).

about the form *of the propositions that list the real requirements of rationality*. If we must invent a new way to talk to express these propositions, that is beside the point.

One of Kolodny's arguments begs the question. He assumes that a requirement of rationality has wide scope only if one can reason one's way into compliance with it from any of the individual states involved in the complex conflict state that constitutes noncompliance with it.[101] But he doesn't argue for this assumption. He just *asserts* that it offers the best way of capturing his 'Rational Response Test', on which a rational requirement governing a conflict of states A and B is wide-scope *only if* "(i) one can rationally resolve the conflict of having A and B by dropping B, and (ii) one can rationally resolve it by dropping A". (2005: 521) I'd reverse his reasoning and view (A) as an obvious counterexample to the assumption. I would not abandon the Rational Response Test. One can appeal to stuff *outside* of a conflict to rationally resolve it. Kolodny would then claim that I am appealing to narrow scope requirements. Well, yes. *Some* requirements are narrow scope—the ones associated with the demand to respond to apparent reasons, in whatever turns out to be the *relevant* sense of 'apparent', which must be *nondoxastic*.

Another of Kolodny's arguments is that wide scope requirements cannot be requirements because they could not be action-guiding. Simply knowing that one violates (A) or (B) does not tell one how to resolve the conflict. But this could not show that (A) and (B) are not requirements. One might have a decisive reason to leave some house. Being aware of this reason as such, one is rationally required to leave. All this fails to determine exactly which door one should use. That does not imply that there is no requirement to leave.

Kolodny's final argument is that nothing could explain the *robust* deontic import of wide scope requirements.[102] But one doesn't have to believe that a requirement in a domain is robustly deontic to believe that it is a requirement in that domain. Arguments of this form couldn't settle what the correct form of any requirement of rationality is. There is no easier explanation of how narrow scope requirements could be robustly deontic, as Kolodny himself argues!

Once we reject Kolodny's first two arguments, nothing stands in the way of a mixed view on which some requirements of rationality are narrow-scope and others are wide-scope. The requirements of intelligibility are wide-scope. The requirements demanding attention to apparent reasons, in a nondoxastic sense of 'apparent' yet to be discovered, are narrow-scope. These requirements tell one *how* to avoid unintelligibility. This is how we can preserve Scanlon's observations even while rejecting all Pure Doxastic Views.

2.3. *The original sin: normative explanation and the conflation of intelligibility and rationality*

---

[101]This is his 'Reasoning Test': "The process-requirement governing the conflict between A and B is wide-scope—that is, one is rationally required (either not to have A, or not to have B)—only if, from a state in which one has conflicting attitudes A and B, (i) one can reason from the content of A to dropping B and (ii) one can reason from the content of B to dropping A" (2005: 521). What a reasonable wide-scoper about a requirement ought to say is that one may only be able to rationally exit the conflict by appealing to factors outside of it. Kolodny has a further assumption that he may use to try to block this—that a requirement only governs a conflict if one can exit the conflict *just* by complying with the requirement. But there are obvious counterexamples to it. All wide-scope requirements do is command us not to be a certain way. They do not tell us *how* to stop being that way. But they are requirements all the same. The point of other narrow scope requirements is to tell us how to get out of the conflicts.

[102]There are other arguments that might seem to cast doubt on the existence of requirements like (A) and (B). Wide-scope requirements are, after all, by themselves symmetrical: one can in principle satisfy them by dropping one's normative belief, or by forming the intention. But the fact that unqualified symmetry is unintuitive is not an argument against wide-scope principles. It is just an argument that there are some *further* requirements of rationality. See Way (2010) and Broome (ms).

It is easy enough to see how people could have missed the problem with the Strong Doxastic View. The problem for the Weak Doxastic View is, however, extremely obvious. Why is it that people sometimes ignore this objection in the recent literature on rationality?

Surely part of it is the plausibility of Scanlon's original observation, which pressures us to accept some belief-involving norms of coherence. But I think the problem goes deeper. Earlier, I said that we can distinguish between *intelligibility* and *rationality* even while granting that rationality is a narrower notion than responsiveness to real reasons. There is a long tradition of assuming that the narrow notion of rationality is precisely the one that plays a role in the distinctive kind of psychological explanation whose aim is to render someone's attitudes and acts intelligible. This tradition is important, since it drew attention to the possibility of separating responsiveness to normative reasons from weaker standings. But it is unhelpful in issuing in the assumption that the weaker standing that plays a role in this kind of psychological explanation is rationality, or at any rate that the *only* place where the ordinary notion of rationality belongs is there.

This problem has its origins in Davidson. Here we can see him identifying a narrower concept by beginning with the kind of "rationalizing explanation" that he famously illuminated:

> The paradox of irrationality springs from what is involved in our most basic ways of describing, understanding, and explaining psychological states and events […]. [T]ake Roger, who intends to pass an examination by memorizing the Koran. This intention must be explained by his desire to pass the examination and his belief that by memorizing the Koran he will enhance his chances of passing the examination. The existence of [motivating] reason explanations of this sort is a built-in aspect of intentions, intentional actions, and many other attitudes and emotions. Such explanations explain by rationalizing: they enable us to see the events or attitudes as reasonable from the point of view of the agent. An aura of rationality, of fitting into a rational pattern, is thus inseparable from these phenomena.
>
> […] Much that is called irrational does not make for paradox […]. Perhaps it is in some sense irrational to believe in astrology, flying saucers, or witches, but such beliefs may have standard explanations if they are based on what their holders take to be the evidence […]. The sort of irrationality that makes for conceptual trouble is not the failure of someone else to believe or feel or do what we deem reasonable, but rather the failure, within a single person, of *coherence* or *consistency* in the pattern of beliefs, attitudes, emotions, intentions and actions. Examples are wishful thinking, acting contrary to one's own best judgment, self-deception, believing something that one holds to be discredited by the weight of the evidence.[103]

This kind of remark was, I suspect, agenda-setting, and when Scanlon provided the impetus for much of the recent literature on rationality by stressing the separability of the narrow notion from questions about genuine reasons, he made similar remarks.[104] The standard list of requirements of rationality that people now address in the literature primarily includes the ones that ban all of the problematic configurations that Davidson lists at the end of the quotation.

But the ordinary, narrow senses of 'rational' and 'irrational' simply are not this narrow. We must break the neat connection between rationality and Davidsonian rationalization, and expand the list. If you have a visual experience as of a cup labeled POISON: DO NOT DRINK, but go ahead and take a sip, you behave irrationally, even if you form no beliefs about the label. This example is *as paradigmatic* a case of irrationality as Scanlon's (perhaps) Davidson-inspired cases. And this is *not* to make a point about genuine normative reasons. If the cup turns out not to have poison in it but rather an elixir that will cure an otherwise terminal illness you have, it surely isn't true that there exists a conclusive reason for you not to drink it. There is a strongly sufficient reason for

---

[103] Davidson (2004: 169–70).
[104] The examples in Scanlon (1998: 25–6) are reminiscent of the ones in the Davidson quotation.

you to drink it, even if you would be irrational in intending to take a drink. So, irrationality can still be narrower than failure to respond to the real normative reasons.

There is a close analogy here with some hypological concepts. The law distinguishes between *recklessness*, *negligence* and *strict liability*. We are strictly liable for anything we voluntarily do that there are conclusive moral reasons not to do. We are negligent for doing what the appearances correctly indicate there to be conclusive moral reasons not to do while ignoring these and thus failing to believe that we are acting wrongly. We are reckless when we *deliberately* disregard the appearances, and so do—at some level or time—acknowledge that we are acting wrongly. The Davidsonian tradition encourages us to think of irrationality as a kind of recklessness. The bad view that conflates rationality and responsiveness to genuine reasons encourages us to think of irrationality as a kind of strict liability. It shouldn't be forgotten that there is an intermediate choice. Irrationality in the ordinary narrow sense resembles this, at least when compared with mere intelligibility on the one hand and responsiveness to genuine reasons on the other.[105]

2.4. *"Apparent reasons" improved: the weak appearance-relative view for ex ante (ir)rationality*

Where should we go from here? Well, I noted at the outset that the most basic and generic observation one glean from cases like Williams's "petrol and tonic" case is that practical rationality turns on how things *appear* from the agent's perspective. More specifically, it turns on the *reasons that appear to exist from the agent's perspectiv*e. In rejecting the Pure Doxastic View, we needn't deny this. We can accept a different incarnation of the Apparent Reasons Theory.

We have a couple options. One is to understand 'apparent' as 'evidence-relative'. On the *Weak Evidence-Relative View*, S is *ex ante* rational in A-ing or intending to A when S's evidence sufficiently supports believing that there are sufficient reasons for her to A. I once argued that this view is preferable to both Pure Doxastic Views, and certain stronger epistemically loaded alternatives like Errol Lord's Accessible Reasons View. But I now would not accept this view.

Why? I think it should be left open whether there is a gap between

> (i) whether an agent's evidence sufficiently supports believing that there are at least sufficient practical reasons to A (or to refrain from A-ing),

and

> (ii) whether the agent would be *ex ante* epistemically rational in believing that there are at least sufficient practical reasons to A (or to refrain from A-ing).

Practical rationality is *very plausibly* parasitic on (ii). But if there is a gap between (i) and (ii), it won't be parasitic on (i). It is plausible that the *quality* of the evidence one has needn't turn wholly on nonfactive mental life, so that subjects in the same nonfactive mental states necessarily have evidence of the same quality. How could X be *good evidence* for P if X's obtaining *does nothing* to raise *objective likelihood* that P? To think it couldn't is the default view. Nothing similar goes for the factors that make for epistemic rationality. This is for reasons like the ones that made us pry practical rationality apart from responsiveness to genuine reasons in the first instance. Epistemic rationality is like practical rationality in being narrow and perspectival. These features require it to supervene on nonfactive mental life., as I argue at length in §3.3. We *want* your nonfactive duplicate in the demon world to be just as epistemically rational. As long as there is an overwhelming *apparent* link between the evidence and the facts, one ought rationally to treat that would-be

---

[105]It should be noted that Scanlon does acknowledge an intermediate notion, which he calls *reasonableness*. Cf. Scanlon (1998: 32–3). But it is even broader than what I would want to regard as rationality in the ordinary sense. Cf. my remarks an earlier note on people who claim to be imposing knowledge norms on practical rationality; as I said there, I think these people are latching onto something like Scanlon's reasonableness.

evidence as if it were good. Since whether the evidence actually sufficiently supports P will turn on how truly good it is at indicating P-like facts, practical rationality can't be made to turn on (i). This is *precisely because* practical rationality goes with epistemic rationality.

An example may clarify the reasoning. Suppose an evil demon temporarily renders Bob unable to distinguish between reality and some convincing daydreams it instills in Bob.106 In the first of these daydreams, Bob is "told" by many different sources that a bottle of water in his fridge has been poisoned. Suppose Bob seamlessly transitions from this daydream to reality. Given his inability to recognize the transition, it is irrational for him to drink from the bottle of water. Yet the fact that he *merely dreamed he had sufficient evidence* for thinking the water was poisoned is not *itself* sufficient evidence for thinking this. More cautiously: it is not *as good* as the evidence he would have if he had received the same experiences while awake, and they were veridical. But *whether or not* his experiences are veridical, he is exactly *as rational* in not drinking from the bottle. This intuition is as robust for me as the original intuitions that supported the division of responsiveness to genuine reasons from rationality in the first place. This is an illustration of why we should not accept an *evidence-relative* view about practical rationality if we make room for views that make the quality of evidence turn on facts about reliable indicatorship. We will end up deeming Bob not *as irrational* as he would be in drinking if the dreams were reality.

So, although we do want it to come out that S is *ex ante* rational in doing whatever act would be sufficiently supported by the total set of reasons *that she can rationally take to exist*, "can rationally take" is not itself going to be understood—at least without a lot of argument—in unqualifiedly evidence-relative terms. Bits of evidence are *good* epistemic reasons, and what makes for the quality of an epistemic reason may well be something beyond a subject's nonfactive mental life.

Going evidence-relative isn't the only way for opponents of Pure Doxastic Views to preserve an Apparent Reasons Theory. We can try to cash out 'apparent' in nondoxastic *and* nonnormative terms. The resulting view would be more neutral than the Weak Evidence-Relative View, and it would, as we'll see, vindicate the idea that practical rationality should track the reasons a person *could rationally take* to exist. Indeed, it will lead to an account of how practical and epistemic rationality are both narrow and perspectival *in exactly the same way*. Let's consider this route.

Many have thought that there are such things as nondoxastic appearances. I am with them. Such appearances show up when one has perceptual experiences, including hallucinatory or illusory ones. It is a familiar fact illustrated by cases like the Müller-Lyer illusion that perceptual appearances can remain even when one knows they are misleading. It is not as if one retains an inclination to believe, or a mere credence in these cases. One's credence that the lines are the same length in a Müller-Lyer illustration could be 100%, and yet the appearance could remain. One should certainly not say, "I'm inclined to some degree to think the lines aren't straight. I just ultimately resist the inclination." *One* is not inclined at all. An *attraction* may be felt. But that would, I believe, be a distinct affair. Being inclined to believe and suffering an attraction to belief are different things. An inclination is still a *stance of* the agent's, while a suffered attraction is a mere *happening to* the agent. All this suggests that some appearances aren't essentially doxastic. And if they are not essentially doxastic, they cannot be *kinds* of doxastic attitudes.107

106If you have novel views about the nature of dreaming (like, say, Ernest Sosa's views), replace this with a different skeptical scenario—say, one where Bob is caused to have hallucinations, and is made unable to tell the difference between the demon-induced hallucinations and his normal veridical experiences.

107It *might* be true that in good, non-illusory cases, appearances can be *partly realized* by beliefs. But there is a huge difference between being able to be *partly realized* by beliefs in good cases, and being *identical* to beliefs. Even if perceptual experiences can, in the good cases, realize the functional role of beliefs—and presumably *other* roles in one's overall mental economy, given their phenomenal character—it hardly follows that perceptual experiences *are beliefs* in good cases in the relevant sense, *pace* Lyons (2009). I will discuss this point more in

One also hosts appearances of this type when one has apparent memories. One can be in a mental state that has the same internal character as an episode of recalling, and yet not actually be recalling anything. This state could persist even once one has learned that there is no fact to be recalled. One also hosts them when one has persisting intuitions, which may be known to be misleading. In these cases, propositions can *strike* one in a certain way even if no inclination to believe or credal investment can be attributed to one personally, as I'll argue at more length in §2.8. I'd venture that for any potential sources of knowledge—perception, memory, intuition, introspection, etc.—one can think of possible appearances that feel just like veridical perceptions, memories, intuitions, introspections, etc., and that could remain even in the face of knowledge to the contrary, and with no inclination to believe or degree of belief.

Exactly what nondoxastic appearances are is a question I can hardly settle, though I'll return to it at significant length in §2.8. For now, what matters is just that there *are* such things as appearances that aren't doxastic attitudes. One can appeal to them in constructing a version of the Apparent Reasons Theory that is neither purely doxastic nor (directly) evidence-relative.

But first, a bit more about the scope of the phenomenon. In addition to appearances that are occurrent, I also think there are *stored* and *dispositional* appearances. Consider my nonfactive mental twin who has been a brain in a vat for nearly his whole life. He can call to mind apparent memories and recollective episodes phenomenally just like mine. The internal processes in virtue of which this is possible might well be identical to the ones in virtue of which I am able genuinely to recall past episodes. Insofar as one has the ability to call episodes like this to mind, one has some stored appearances. These should not be confused with *dispositional appearances*. Maybe you've never considered the prospect of stabbing yourself in the hand. If this did occur to you, it would probably strike you as something there are conclusive reasons not to do. The appearance needn't be doxastic. Perhaps you'll fully convince yourself that you must do it, so that *you*, at the person level, won't be *inclined to think* there are reasons not to do it. No *investment* in the correctness of the appearance, *even partial*, would be attributable to you. But *it* would still *strike* you as something there are reasons not to do.108 This is a dispositional appearance.

One last distinction. The appearances we've discussed so far are *prima facie*. *Prima facie* appearances can conflict. Take Locke's old case of perceptual relativity, where, after sticking one hand in hot water and another in cold water, one sticks both into the same pool. One perceptual appearance presents the pool as hot, the other as cold. These are *prima facie*. There is also such a thing as an *ultima facie* appearance, which is revealed by Locke's case. The appearances in his case plausibly *neutralize* each other, so that one ends up *pulled* toward the content of neither, relative to the *totality*

---

§2.8 when I discuss the nature of appearance at greater length, and clarify how an Weak Appearance-Relative view of rationality does not face the same objections as the Weak Doxastic View.

108The italics here indicate an important difference. Some may say that in this case, *one is inclined to believe that there are conclusive reasons not to act* but that one simply resists one's first-pass inclination. See, e.g., Williamson (2007: 217). But there is an obvious and key structural difference, I believe, between *something's striking you* a certain way—in an important sense of this phrase—and *your being inclined on first pass to see it that way*. The inclination is attributable to you, at the person-level, even if it isn't *ultimately endorsed*. The striking is not felt as being attributable at the person-level in the same way. This is reflected in language: contrast "*it* strikes *you* a certain way" with "*you* take *it* in a certain way". This is also part of the phenomenology in some cases, and indicates a similarity between perceptual and intuitive appearance. Strikings are felt as coming "from the outside". Inclinations aren't. They are attributable to one. I'll elaborate on these thoughts in §2.8.

　　Now, I agree that some intuitions are mere inclinations to believe, and some are beliefs. That is because ordinary talk and thought about intuition is disjunctive, which is one reason why I myself tend to avoid disputes about intuitions. There is no one phenomenon worth discussing, and none that actually figures centrally in philosophical reasoning. I'm interested in one disjunct, and I think it is wrongheaded to ignore the linguistic and mental phenomenology and pretend that there are no deep differences.

of experience. So the water doesn't *ultimate strike* one as hot or cold. There is simply no way in which it ultimately appears. The same distinction can be drawn for other appearances.

I'll speak of the *total ultimate appearance* for S as the *ultima facie* appearance that can be constructed from all stored, occurrent and dispositional appearances of S's at a given time. Hence, the total ultimate appearance for S at *t* has it that P only if the following is true: other things being equal, if all stored and dispositional appearances existing at *t* were to be made occurrent, the *ultima facie* appearance would have a content conditional on which P is true or likely enough to be true. S's total ultimate appearance is a perspective on the world created by the totality of S's nonfactive mental life, a good bit of which need not be conscious. It is the one relevant for understanding the perspectival character of rationality. To start to see the view I like about *ex ante* (ir)rationality, consider this *Weak Appearance Relative View*:

> **WARea**: Setting aside all conflicts with other paradigmatically motivational attitudes and assuming satisfaction of all wide-scope requirements of coherence,
>
> > S would be *ex ante* rational in intending to φ (or in φ-ing) iff, according to the total ultimate appearance for S, there are (at least) sufficient reasons for φ-ing,
>
> and
>
> > S would be at least somewhat *ex ante* irrational in intending to φ (or in φ-ing) iff, according to the total ultimate appearance for S, there are sufficient reasons not to φ.

This view is like Scanlon's Weak Doxastic View in building explicitly normative contents (hence the name), but unlike the Weak Doxastic View in not being a Pure Doxastic View.

This is not *quite* the view we want. This is because the *strength of the total ultimate appearance* must be taken into consideration, and not just *what* that appearance paints. Suppose—to take an unrealistically simplified case—that there are only two moral intuitions that could arise for S that are relevant to answering the question of whether A-ing is right. One is a tiny bit stronger than the other, and favors a positive answer to this question. If A-ing is right, there are sufficient reasons for A-ing. So, sufficient reasons for A-ing do ultimately nondoxastically appear to exist. But the appearance is very weak. This makes it less plausible to come down on the verdict that S would be (*ex ante*) rational in A-ing. So the *strength* of the appearance needs to be taken into consideration: how rational S would be (*ex ante*) in A-ing should turn both on the strength of the reasons that ultimately nondoxastically appear to exist, and on the strength of the appearances.

Exactly how the second proportionality works is unclear. There may well be no fact of the matter about the exact contours it takes. This is no surprise: rationality is plausibly a vague concept. Still, that there should be such a proportionality is plausible. So we can revise WARea:

> **WARea***: Setting aside all conflicts with other paradigmatically motivational attitudes and assuming satisfaction of all wide-scope requirements of coherence,
>
> > S would be *ex ante* rational in intending to φ (or in φ-ing) iff, according to the total   ultimate appearance for S, there are (at least) sufficient reasons for φ-ing,
>
> and
>
> > S would be at least somewhat *ex ante* irrational in intending to φ (or in φ-ing) iff, according to the total ultimate appearance for S, there are sufficient reasons not to φ,
>
> where degrees of (ir)rationality are proportional to the strength of the ultimate appearance.

Similar proportionality theses should be upheld for rational requirements and not just permissions. When total ultimately apparent reasons apparently decisively support φ-ing, it will not immediately

follow that S is rationally required to φ. That only follows if this appearance is sufficiently strong. If it is very weak, then S would not be fairly charged with the flat-out, harsh criticism of irrationality for deciding not to φ: S need only be irrational to a very minor degree.

Even this view will call for a significant revision, as we'll see in a bit. I'll assume it for now.

2.5.   *Ex post (ir)rationality*

I've only considered how the Weak Appearance-Relative View should apply to *ex ante* evaluations. The *ex post* story is more complicated. One natural idea is that we only regard S as rational in A-ing *ex post* if S acted for the considerations that ultimately nondoxastically appeared to be sufficient reasons from S's perspective. Moreover, S needs to act for these reasons in the dominant sense— S needs to *comply* with them and not just *conform* to their dictates. This was what I called the *strong sense* of acting for a reason §2.2 of Chapter 1. So, we want something like

> **WAR1ep**: Setting aside all conflicts with other paradigmatically motivational attitudes and assuming satisfaction of all wide-scope requirements of coherence, S is *ex post* rational in φ-ing iff (i) it is *ex ante* rational for S to φ, and (ii) S does φ for   the apparent reasons in virtue of which it is *ex ante* rational *in the strong sense*.

This view plausibly allows that S can continue to act for the same reasons whether or not she is in a good case. The difference between good and bad cases—i.e., cases where appearances are veridical and cases where they are misleading—is not found in the motivating reasons for which S acts, but rather in their quality as normative reasons. In the bad case, the appearance does not come in as a new reason for which the agent acts, but as a background condition in virtue of which the agent's acting for the same old reason remains rationally impeccable.

What changes from the good case to the bad case is in the first instance just whether the propositions that are S's reasons for A-ing are *true*. Of course, this does involve giving up on the factivity of the 'that'-clause in motivating reasons claims like "S's reason for A-ing was that P". But this, I think, is not a bad result. It is the truth. It is odd to think S's own reasons for acting could change without change in S's mental life, with S clueless to the change. We would have no choice but to accept this if we bought the factivity of motivating reasons ascriptions. The intuitions that support factivity simply mistake pragmatics for semantics. If, in describing S, we didn't qualify our description of the propositions that were her reasons for acting with something like "which was how it seemed to her then", we will *implicate* that her reason was a fact and not a proposition that—by the way—seemed to be true. But given that it is clearly wrong to say that S's own motivating reasons change, and do so totally unbeknownst to her, we can clearly see how the implicature can and ought to be canceled. As Dancy notes, we can do this:

> A perfectly correct explanation of an act might be:
>
> His reason for doing it was that it would increase his pension.
>
> But such an explanation cannot be factive […] since it can perfectly well be expanded thus:
>
> His reason for doing it was that it would increase his pension, but sadly he was   mistaken about that.[109]

My semantic intuitions are with Dancy's. If one tells the right story, claims like this one can sound

---

[109] Dancy (2004b: 25−6).

perfectly acceptable. When they sound bad, that is for pragmatic reasons.[110] This may appear to contravene the Davidsonian idea that rationalizing explanation is a species of causal explanation in the most straightforward sense. After all, causal explanations are factive. But as Dancy suggests, the two stories can go together. Often the way in which someone has P as her motivating reason is by believing P or recognizing P's truth (in the good case), with the belief or recognition playing, together with paradigmatically motivational states like desire, its causal role. This causal story is about *how it is* that someone acts for some reason. It is the *background* to the kind of normative explanation that lets us in on what the agent saw what she was doing, and illuminates how it could be true that when the case is good, someone can act for a good reason.

This does force us to clarify what could be meant by acting for an *apparent* reason in the strong sense. In Chapter 1, the idea was that S does A for a good reason in the strong sense when *what makes that reason a good one for A-ing explains why it is S's reason for A-ing.* This was intended to be understood like so: the facts that make the reason a good reason have to be tracked by the agent and *thereby* be playing an explanatory role—as I said, *via* a practical reasons-sensitive ability of the agent's that ensures a mirroring between her motivational attitudes and the objective normative structure of the world. *That* cannot be the story here, since the reason will no longer be a good reason when the proposition that constitutes it is false. The old normative structure *disappears.* A non-obtaining state of affairs cannot play the kind of explanatory role that we had wanted.

So the *appearance* that there is a good reason will play the explanatory role. We introduce the

> **Secondary Strong Sense**: S does A for an apparent reason R in the *secondary strong sense* iff the appearance that R is a sufficiently good reason is what explains why the propositions that constitute R are the agent's own reasons for A-ing.

Like I suggested in Chapter 1, the *explanation* in this claim ought to proceed by way of a *reasons-sensitive ability:* it is just that it is now an *apparent reasons*-sensitive ability, rather than a *real reasons-sensitive* ability. This will be a disposition to heed and act for a range of apparent sufficient reasons sufficiently like R in just the circumstances where, if the appearances were veridical, the reasons would be sufficiently good. The disposition should also be one *not to act* just for these reasons in circumstances where, if the appearances were veridical, these reasons would not be sufficiently good. R's being S's reason for φ-ing is explained by R's apparent goodness in the sense relevant to this claim when *S's φ-ing for R manifests such an apparent reasons-sensitive ability.* So the "strong sense"

---

[110]One can imagine agents coherently agreeing. Consider the following case. It is 3:55. Bill asks Jane what her reason for leaving is. Jane says: "The talk is at 4! I need to run!" Bill tape-records this. At 4:02, Bill shows up where Jane thought the talk was to be held. There is no talk. Bill asks Jane again for her reason for being there, but simply answers his own question by playing the tape recording. He says, "So! Your reason was that the talk was at 4. Sadly, you were mistaken." Jane may then coherently say: "Well, yes. I wish I had known better. Let me explain why I didn't know better, so you can see that I acted rationally."

This case also reveals why asserting that one's reason for A-ing was that P when P is false can sound unacceptable. Any agent who learns that P is false will obviously want to change her mind about acting on the assumption that P. It would be clearly irrational for her not to do so, given what she now knows. Because of this, it would be bizarre if, having learned that ~P, an agent said *without any qualification* that her reason for A-ing was that P. That's because this would represent her as being in an obviously incoherent or negligent state. So she will typically go on to say: "Well, I did *believe* that P. But look: I had good reasons for that." This is not her indicating what her reasons *really were* from the get-go, but simply her excusing herself for the mistake of acting on a false assumption. In our case, we can sidestep all the pragmatics, because the interlocutor has already set things up so that the agent can go on to add the qualifications in such a way that they clearly are just serving as *excuses*, rather than as indicating what the original motivating reasons *really* were. When she does this, there is nothing problematic about saying that, yes, her reason for A-ing was P, and sadly she was mistaken.

Thanks to Errol Lord for pressing me on this issue, and for prompting me to create the example.

in WAR1ep must be understood as the secondary strong sense.

What about *ex post* irrationality? Unsurprisingly, the story is going to be a bit more complicated than the story about *ex ante* irrationality. There are *two* ways in which an agent can be irrational *ex post*. One way is for her to do what she would be *ex ante* irrational in doing. Another way is for her to fail to act for the apparent reasons that make it *ex ante* irrational for her so to act. So:

> **WAR2ep**: Setting aside all conflicts with other paradigmatic motivational attitudes and assuming satisfaction of all wide-scope requirements of coherence, S is at least somewhat *ex post* irrational in φ-ing iff either (i) the reasons that ultimately nondoxastically appear to exist sufficiently support not φ-ing, or (ii) S fails to act for the apparent reasons that make it *ex ante* rational to φ in the secondary strong sense.

As before, caveats are in order about the effects of the *strength* of the ultimate nondoxastic appearance on S's degree of (ir)rationality. Since no new caveats are needed beyond the ones detailed for *ex ante* rationality, I'll just take what I said before as given to save words.

2.6. *Have we avoided the problem of difficulty? Recognizing the primacy of* ex post *(ir)rationality*

Our account shares with Scanlon's the feature of building *explicitly normative contents* into the mental states that determine what it is *ex ante* rational to do. If there merely ultimately nondoxastically appear to be nonnormative facts that would give sufficient normative reasons for A-ing if they obtained, nothing follows about whether S would be rational or irrational in A-ing. Something only *follows* if we know about how the total nondoxastic appearance *normatively represents* the facts. This avoids the problem that plagued the Weak Doxastic View. I think our view also avoids the easiness that plagued the Strong Doxastic View, though I will get into this more in §2.8. The appearances on our view function as *external constraints* on one's doxastic state. Avoiding the belief that one has decisive reasons to A won't allow one to avoid a rational calling to A. The appearances that concern us run deep, and well beyond belief or credence.

But one might worry that, for this reason, the particular problem of difficulty for the Strong Doxastic View has simply been exchanged for a different one. Notice that how the total ultimate appearance normatively represents the facts will be a product of the balance of many normative intuitions, including dispositional and stored ones, as well as other factors that may be occurrent, like normatively loaded perceptual experiences and certain emotions.[111] The result is intricate in something like the way the product of reflective equilibrium would be. Obviously, successfully achieving reflective equilibrium is hard. So it looks like a problem of difficulty may remain.

Indeed, remember how in objecting to the Strong Doxastic View we considered someone at an intermediate stage in her normative reflection, where she was unsettled between, say, rule utilitarianism and a modest deontology with prerogatives. These theories may strongly disagree about the deontic status of an act-type A. If S knows that X is an A-type act, the Strong Doxastic View will imply that she would only be rational in obeying the verdict of the *true theory* in deciding whether to X. Something close to this would seem to hold on WARea*, if we include stored and dispositional appearances within the base of appearances that constitute how things ultimately nondoxastically appear. It might take *a lot* of reflection to see how the appearances come down on these theories. If they come down in favor of rule utilitarianism, would it be *irrational ex ante* for S to fail to do what it requires when the deontological view permits one to do otherwise? That

---

[111] The idea of normative perception is not as wild as one might think if one's views about perceptual content are *already* liberal—say, as liberal as Bayne (2009) suggests. See Cullison (2009) and McGrath (2004) for an illuminating discussion of this topic. In §2.8, I'll invoke a commonsense version of this idea.

is very implausible. How can we resolve this problem?

There are many options. One thing we could do is confine focus to *occurrent* appearances. But this seems too permissive. In S's less careful or thoughtless moments there may be no occurrent appearance representing an object of choice in any normative light. That should not suggest that there is no fact of the matter about whether it would be rational *ex ante* for S to choose it.

We could also try to build in restrictions on the ease with which S could call to mind stored and dispositional appearances, and weigh them against each other. We could then add some scaling factors by which the degrees of rationality that our theory currently yields are multiplied: the harder it is to call appearances to mind, the less effect these appearances have on final degrees of rationality. But this feels *ad hoc*. One craves an explanation for *why* we need the scaling factors apart from the fact that it merely helps us to capture the extensional intuitions more extensively.

I think it is better to take a different route, one that puts emphasis on a core fact about rationality we observed at the outset in §1.2 of this chapter—viz., the *act-agent blur* that is distinctive of evaluations of rationality. The existence of act-agent blur speaks in favor of taking *ex post* rationality to be the more central of the two notions. *Ex post* rationality is even more clearly focused on the agent than *ex ante* rationality. We need to know whether the agent's ability to respond to apparent reasons in the secondary strong sense was exercised in order to know whether she acted rationally or irrationally in this sense. Now, in Chapter 1, I did object to Turri's suggestion that we ought to explain propositional *justification*, which is an *ex ante* matter, in terms of doxastic *justification*, which is *ex post*. I'll remain resistant until there are convincing reasons for thinking that rationality and justification are more closely tied than general considerations suggest. But I am not opposed to putting *ex post* ideas first in all cases. With rationality—an idea for which there is act-agent blur—it seems like the right thing to do.

What would this suggestion involve? Notice that if an agent simply lacks the capacity to see what the total ultimately apparent reasons are, this agent will not be in a position to act for these reasons in the secondary strong sense. Their apparent quality could not play the explanatory role it needs to play for an agent like this to act for these reasons in this sense. If we redefined *ex ante* rationality to track whether an agent is *in a position* to act in a way that would render her *ex post* rational, we could capture this fact, and thereby avoid the problem of difficulty.

So what we need to do is add another revision to WARea*:

> **WARea****: Setting aside all conflicts with other paradigmatically motivational attitudes and assuming satisfaction of all wide-scope requirements of coherence,
>
> > S would be *ex ante* rational in intending to φ (or in φ-ing) iff
> >
> > > (i)   according to the total ultimate appearance for S, there are (at least) sufficient reasons for φ-ing, and
> > >
> > > (ii)   S is in a position to respond to these in the secondary strong sense,
> >
> > and
> >
> > S would be at least somewhat *ex ante* irrational in intending to φ (or in φ-ing) iff
> >
> > > (iii) according to the total ultimate appearance for S, there are (at least) sufficient reasons against φ-ing, and
> > >
> > > (iv)   S is in a position to respond to these in the secondary strong sense,
> >
> > where degrees of (ir)rationality are proportional to the strength of the ultimate appearance.

We aren't exactly reversing the definitional order here. *Ex post* rationality makes no explicit appearance in these definitions. But the addenda only make sense if we understand that *ex post*

rationality is the more fundamental thing. *Ex ante* rationality turns out on this account to amount to being in a position to act rationally in the *ex post* sense. This is analogous to Turri's reversal of the familiar priority order of propositional and doxastic justification, though we should not expect *his* reversal unless there's an argument for act/agent blur for the deontic in epistemology—something that would holds *only if* one can connect justification and rationality in epistemology in a way in which we cannot plausibly do in the practical sphere.

2.7.   *Overintellectualization?*

I turn to a different concern about the Weak Appearance-Relative View. One might worry that there is implausible overintellectualization in it—specifically, in the demand that there be normative content in someone's nondoxastic appearances for assessments to be made of her rationality. One might say: "For it to appear (even nondoxastically) *that* there are sufficient reasons not to A, one must have the concept of a reason. Yet children and sophisticated animals lack the concept of a reason. Surely they can be rational or irrational all the same."

I am generally unconvinced by overintellectualization charges of this type. Often when I hear objections like this, it is really the objector who is doing the overintellectualizing. That is true here. Why think children and sophisticated animals cannot have the concept of a reason? It cannot be that children and such animals lack the ability to *articulate* normative thoughts. Concept possession does not require articulability. It also can't be that children and animals lack sophisticated *normative theories*. Having concepts within a certain domain does not require one to be able to formulate sophisticated theories about the things in that domain. But this is all that is *relevantly clear* about children and higher animals. They can't articulate any normative thoughts, and they also lack normative theories. If having a concept requires neither of these things, I don't see why it is so clear that children and sophisticated animals lack the concept of a reason.

Let me be a bit more defensive. On a plausible kind of view about concept possession, to have a concept of something *as F* involves having *a way of regarding something as F*, and to have a way of regarding something as F is to have certain dispositions to act or think, given the presence of certain further beliefs or nondoxastic appearances.112 Having the concept of a reason could simply amount to being able to regard something as a reason. Being able to regard something as a reason could in turn amount to having dispositions to act or think in certain ways, *given certain further nonnormative beliefs or nondoxastic appearances*. Now I take it that the objector is not assuming that children and sophisticated animals cannot have nonnormative beliefs or suffer nondoxastic appearances. To understand the behavior of children and sophisticated animals in a way that would make it at all clear that they are rational beings in the first place, they will have to be understood in this way. But the very same considerations that make this intelligible also make it perfectly unclear why children and sophisticated animals can't have the concept of a reason! At any rate, that is clear on a natural and not overly intellectual account of concept possession.

What exactly might the dispositions be? The style of answer is straightforward. To regard *that P* as conclusive reason not to A is, at minimum, to be disposed to refrain from A-ing given an *ultima facie* awareness (doxastic or non-) that P. To regard *that P* as a conclusive reason for A-ing is to be

---

112Note that I am not saying that having a concept *is* having certain dispositions to act or think. I am just saying that it *involves* this, as a point of necessity. I thus do not prejudge the important dispute between the Representational Theory of Mind as defended (e.g.) by Fodor and various forms of "concept pragmatism". A friend of the Representational Theory of Mind can (and should, I think) hold that having certain dispositions to token a symbol in the Language of Thought is a *sustaining mechanism* by means of which that symbol gets the content it has, and even hold that a symbol in LOT can't have the content it does without being used in certain ways. Cf. Margolis (1998) and the Introduction in Laurence and Margolis (1999).

disposed to A, given an *ultima facie* awareness that P. Similar if slightly more complicated stories can be told for regarding considerations as having different positive or negative normative weights. The stories will simply involve somewhat more intricate dispositions to act or think in the light of somewhat more intricate nonnormative beliefs. Insofar as a creature has systematic dispositions of these kinds, it will be as plausible to view the creature as having normative beliefs *vis-à-vis* the bearing of P on A-ing as it is to treat the creature as having nonnormative beliefs. Insofar as the objector isn't assuming what *would* be an overly intellectual view—namely, that children and animals cannot have *nonnormative beliefs*—there seems to be little extra trouble in seeing how they could have normative beliefs. Obviously, this doesn't mean that they must be able to articulate them, or form sophisticated theories about reasons for action. That's a completely general point that simply shows what it *doesn't take to have concepts or beliefs*.

Once one has the right account of having concepts, beliefs and concept-involving nondoxastic appearances, one will have the resources to make it plausible that children and higher animals can have normative concepts, beliefs and concept-involving nondoxastic appearances. The overintellectualization lies in the assumptions that the objector must implicitly be making about what it takes to have concepts and beliefs. Once these assumptions are rejected, there will be no *special* difficulty in attributing normative concepts and beliefs.

2.8.   *Have we avoided the problem of easiness? More on the nature of appearance*

There is a final concern one might have about the Weak Appearance-Relative View—namely, that it doesn't really avoid the core problem of easiness that plagued the Weak Doxastic View.

To bring out the best way of stating this concern, recall the details of the original objection to the Weak Doxastic View. Recall that, on that view, one could avoid practical irrationality just by changing or refusing to change one's normative beliefs. Also recall that this was just the *beginning* of the objection. The real objection was then, in my earlier words, that "straightforwardly false and crazy normative beliefs are easy to maintain if one is sufficiently dogmatic or thick-skulled in ignoring intuitively obvious normative truths". When put this way, the Weak Appearance-Relative View seems to avoid the objection: after all, it makes practical rationality turn exactly on what is more or less obviously a reason from the point of view of the total ultimate nondoxastic appearance. That, we said, is what makes normative beliefs *ex ante* rational, all things considered.

The best way to put the remaining worry is that we have only deferred the objection. There are two ways of pressing this thought. To see the first, note that the total ultimate nondoxastic appearance *could* be impoverished with respect to what it implies about the reasons that exist, and *in a way that causally depends on defective beliefs or other attitudes*. This may look problematic. Couldn't a murderer be deprived of the ability to be struck with the intuitions that contravene her deviant moral beliefs, given the sheer strength of these thick-skulled (and thick-skinned) beliefs and her desire to murder? And wouldn't she still not be fully rational, even in the familiar narrow sense, for acting on these beliefs? Indeed, for any potential counterexample to the Weak Doxastic View, it seems we can imagine an extreme variation where a person's dogmatic or thick-skulled beliefs causally deprive her of the contrary ultimate appearances. Aren't these counterexamples?

I agree that such cases are possible, though they are, rather crucially, far more extreme and less numerous than the objector lets on.113 But I deny that they are relevantly analogous to the

---

113Why? Note that even if it doesn't *prima facie* strike some sociopath, say, that infanticide is usually wrong, this person will often have a ton of clear apparent evidence available *that almost everyone regards infanticide as wrong*. So it will still *ultima facie* appear that infanticide is usually wrong: the truth of this claim is *likely*, *conditional on all the extra clear apparent evidence*. So it is much harder to imagine the putative counterexamples than it might have seemed, given the schematic objection. I don't find it remotely clear what we ought to say about the

counterexamples to the Weak Doxastic View.   So I deny that they are counterexamples to the Weak Appearance-Relative View. To see why I say this, I'll draw on an analogy with perception.

One's expectations can causally impact what perceptually appears to one to be the case. A perceptual appearance that causally owes to certain expectations or beliefs isn't thereby rendered a doxastic appearance. Consider aspect-switching cases, as with Necker's Cube. Given the obvious phenomenal difference between the two ways of seeing the cube, the difference in ways of seeing does not *consist in* a difference in doxastic attitudes. The expectations could differ without the relevant phenomenal difference being *exhausted by* this difference.

Now, one can imagine someone who has been primed to see the figure in only one way, for merely explanatory albeit cognitive reasons she can't detect now. Perhaps, in the past, she saw lots of actual cubes oriented only in this way, but now cannot access this information. Suppose this person's expectations are so deeply entrenched that they make it impossible for her to aspect-switch the figure into the other orientation. And suppose that upon seeing Necker's Cube for the first time ever, this person is asked in an innocent-seeming way: "What orientation do you think the represented cube was intended to have?" Even if the obscured background expectations that *cause* our subject to see only one orientation were defective—perhaps she was told when she saw the earlier cubes that the lighting was tricky, and irresponsibly ignored this—it could still intuitively be *rational* for her to give the answer that fits with how things perceptually appear now. What *else* would be rational? Saying "I dunno", and just suspending judgment? Not if the question was asked so that it didn't reveal that the answer was not the obvious one, and if our subject was really unfamiliar with aspect-switching, and if she forgot what made the earlier bases for her hidden expectations defective.   And *nota bene*: if one of those things weren't true, the total ultimate nondoxastic appearance for her wouldn't require the naïve answer.

That is not to say that she would be justified in believing what the ultimate appearances now suggest. The Weak Appearance-Relative View was never intended to be a theory of justification. Assessments of rationality do not turn on facts about the etiology of appearances in the way that facts about justification may. This is true unless the facts about etiology were themselves clearly accessible to the person at the time—in which case the person would be irrational in a way that is perfectly consistent with our view. What this kind of case really shows is not that a Weak Appearance-Relative View about rationality is false, but rather that the deontic significance of rationality can be thrown into question. It can be thrown into question *precisely because* it can require one to adopt beliefs that are highly defective on other clearly deontically significant dimensions of evaluation, such as justification. That, at any rate, is what I'll argue in Chapter 3, and *partly on the basis of cases like this*. The lesson, if there is one, is not about the nature of rationality or its dependence on etiology, but rather about its questionable deontic significance.

If perceptual appearances and intellectual appearances (some intuitions) are things of the same kind, this parallels what we ought to say in the cases used to generate the residual worry. It *is* true that for every case where someone's beliefs are impoverished, we can find another possible case— a far more extreme one, as I noted—where a counterpart's ultimate nondoxastic appearances are also impoverished, and in a way that *causally owes to the defective beliefs*. That can't be the objection. The objection to the Weak Doxastic View was not that someone's beliefs can be impoverished or epistemically unjustified or generically defective. It was rather that her beliefs can be *obviously at odds with how things ultimately appear*, and that acting on the basis of beliefs of this kind is practically irrational *partly because the beliefs would then be irrational*. Given the perceptual analogy, no similar complaint arises for the Weak Appearance-Relative View. The people in the cases that generate the residual worry *are* rational in believing what they believe at the later time, though they are often

---

agents' rationality in these far-out cases. This makes them harder to use as counterexamples.

defective in other ways, and may have been irrational *in the past*. The ultimate court of appeals for rationality is total ultimate nondoxastic appearance. While these people's ultimate appearances may sometimes be incapable of transmitting sufficient justification to their beliefs and acts, there is a huge difference between that claim and the claim that ultimate appearances can be *irrational*, so that acts or attitudes formed in the light of them would have an *inherited irrationality*. That isn't true. And *that's* the relevant disanalogy.

The real objection to the Weak Doxastic View is that dogmatic or thick-skulled normative beliefs can be *irrational*, and that this irrationality plausibly does infect the people's irrationality in acting on these beliefs. Appearances cannot be irrational, and if nothing in the sphere of total ultimate nondoxastic appearance opposes a normative belief, it cannot be *ex ante* irrational. Ultimate appearances simply are the final court of appeals within the sphere of evaluations of rationality. To go any *farther* would be to change the subject, and to give up on core constitutive facts about rationality—e.g., its perspectival character and its narrowness. To go any *less far* would be to open the view up to the *real* problem of "easiness" that was originally pressed against the Weak Doxastic View, and to collapse the distinction between mere intelligibility and rationality.

The real lesson is that while rationality does bottom out in responsiveness to nondoxastic appearances which are not themselves evaluable on the dimension of rationality, the attitudes or acts that are formed or done in response to the apparent reasons may still be evaluated in a different way that does recognize etiology. The facts about the etiology of the appearances do not show that the appearances can be irrational or even unjustified, though one could call them "*defective*" in a broad, purely evaluative sense, as one could call many things (e.g., books). They are not themselves motivated by reasons or formed for reasons. Of course, there are *explanatory reasons* for their existence. These might even be of a type that could motivate other states. But that is still very different. The reason why I'm in pain might be that I think I made a mistake. The belief made me want to hit myself. I did. Still, my pain is hardly motivated by this belief.

So, there is a real disanalogy. We can accommodate the residual intuitions without giving up on the Weak Appearance-Relative View of rationality. At any rate, that is true *if* nondoxastic perceptual appearances and the intuitions that matter on our view are relevantly similar states. The point used to create the analogy was that nondoxastic perceptual appearances aren't the sorts of things can one can have *for (apparent) reasons*, though they can be defective in other ways. Together with some other facts about their nature, this enables them to serve as regress-stoppers *within the sphere of evaluations of rationality and irrationality*, even if evaluations of rationality and irrationality aren't the only evaluations, and turn out, indeed, to be deontically insignificant. If intuitions of the kind appealed to by the Weak Appearance-Relative View are similar—as I've assumed they are—we would have a response to the residual concern by analogy.

It is at this point that the second version of the worry comes up. Many have resisted the thought that intuitions ought to be viewed as relevantly like perceptual experiences. Many have claimed instead that intuitions are more like beliefs. If that were true, there might be a serious concern like the one for the Weak Doxastic View, at least for our story about practical rationality.

Before replying, I should issue a disclaimer. I do not think that there is a single or common phenomenon under the heading of "intuition" that does the work for philosophical methodology. While I do think there is a type of intuition that is more like perceptual experience than belief, I do not think it is the only state worth calling "intuition". The Weak Appearance-Relative View does not require saying otherwise. *Some* intuitions are inclinations to believe or even credences with a certain etiology. These are properly relied on by philosophers when competently formed. Nothing I will be saying here is intended to rule out this claim about propriety in philosophical methodology, or is intended to presuppose that everything picked out by the highly flexible word "intuition" is more like perceptual experience than belief.

What I will say is that I think there is an important class of states that can be called "intuitions" (along with other states) that are more like perceptual experiences than beliefs. Of course, most defenders of the competing view wouldn't claim that all intuitions just are beliefs. The existence of intuitive "illusions", well exemplified by cases of paradox, leads most people to concede otherwise. The competing view I have in mind is one on which all intuitions are either *inclinations to believe* or *degrees of belief* with a certain ancestry, which may or may not, to use Sosa's nifty word, "graduate" into full-fledged beliefs. The inclinations on the view at issue are *prima facie*, like the *attractions to belief* Sosa (2007: 49) mentions here: "*Prima facie* seemings are *relative*: relative to the look of the Müller-Lyer lines, for example, one is *prima facie* attracted to think them incongruent, while the result of a measurement attracts one to the opposite conclusion."

Why think some intuitions aren't beliefs, and are more like perceptual appearances? I'll start with three general differences, and then explain briefly how I'm thinking of some *normative* intuitions, which play an obviously key role in the Weak Appearance-Relative View of practical rationality.

One difference is simply with respect to *person-level attributability* and *evaluability* on the dimension of rationality. To bring out what I mean, we can consider the perceptual case again. Of course, as I said before, I agree that perceptual experiences can be evaluable as *defective*, and that a person can sometimes be defective in a broad sense for relying on them. What isn't true is that one can be rational or irrational for having perceptual experiences, or that they are attributable to one *qua* cognitive agent. In this respect, they differ from credences and inclinations to believe. One can be irrational for having credences, and partly because these are attributable to one *qua* cognitive agent. Once you've learned that the Müller-Lyer lines are congruent, you ought rationally to assign credence 0 to the hypothesis that they are incongruent. But you still perceptually experience them *as incongruent*.114 You are not irrational for that. It is also wrong to say that *one* is still *inclined* or *tempted to some degree* to think them incongruent, having learning that they aren't. *I* am not inclined to think so at all. If I said, "Yes, I'm inclined to think them incongruent, but just resist this inclination of mine", I would speak either irrationally or falsely. *I* have no inclination to be resisted. Thankfully: otherwise I would be irrational. All the same, I see them as incongruent. That isn't the sort of thing for which *I* can be rationally criticized. This is partly because it isn't something attributable to me—a *stance of mine*—like an inclination is. If I were built *better* as a perceptual device, I wouldn't have the experience. But that is a different criticism.

I would say exactly parallel things about some intuitions. Consider Galileo's paradox, which I find a more striking example than the one that often gets discussed (viz., Naïve Comprehension). Consider the positive integers: 1, 2, 3, 4…. And consider their squares: 1, 4, 9, 16…. Before learning about transfinite arithmetic after Cantor, it is intuitive that there should be more positive integers than squares of them. After all, some positive integers are perfect squares, but many are not. But the subtler mind knows that these sets can be put into 1-1 correspondence, and that infinite sets have the same cardinality when their members can be put into 1-1 correspondence. So the size of the set of positive integers is the same as the size of the set of their squares. Now

---

114Here I would resist Sosa (2007: Ch.3, n.3)'s claim that the lines aren't really incongruent in experience: "The lines on the page are of course equal in length. How about the lines in the mental visual field? The lines in one's image do not seem to change in length as congruent parallel lines are inserted sequentially between the two […]. The lines *in one's visual field* are hence arguably congruent even before the lines are inserted." Sosa is there, I think, relying on an impoverished conception of perceptual experience, and focusing too much on the low-level phenomenal content of the experience. Consider, to see why I say this, the famous example of a coin held at an angle. I would opt for a view of perceptual experience on which the coin really is seen as circular. But, of course, if you held up a transparent outline of an oval, you could make it align with the "visual image" of the coin. That simply shows that the phenomenal content of experience goes beyond low-level information. I'd say the same about the Müller-Lyer lines. Cf. Boghossian (2009).

that I know that fact, I assign credence 0 to the original hypothesis that stuck me. *Pace* Williamson (2007: 217), it is strained to say that I am still *inclined to think otherwise*, and just ultimately resist my dumb inclination. No inclination is attributable to me that needs to be resisted, any more than in the perceptual case, with the Müller-Lyer lines. Nevertheless, I can still feel Galileo's intuition. The result is still *striking*. I am not less than fully rational for that.

We can be rationally criticizable for both credences and inclinations to believe when we have clear knowledge to the contrary. We cannot be rationally criticizable *for being struck* with an intuition when we have clear knowledge to the contrary, at least in some cases, like this one. So some intuitions are neither credences nor inclinations to believe. This is analogous to the argument for the same kind of conclusion in the perceptual case (i.e., with Müller-Lyer lines).115

This is one salient analogy between perceptual experiences and some intuitions. Another one is hinted at by the paradigm locution for indicating the presence of the kind of intuition that interests me. There is a clear difference between the kind of state that we report with

> (1)   S is struck by (the fact that) P,

and the kind of state we report with

> (2)   S is inclined to think that P.

If I say something like (2), replacing "S is" with "I am", I invest myself to some degree in P's truth. Nothing like this necessarily shows up when I say something like (1)—particularly in cases where I am thinking about something for the first time. In those cases, it is perfectly acceptable for me to say: "You know, it strikes me that P. But I really have no clue what to think about P." Indeed, in some of these cases, I may not even be able to precisely identify *what* is striking me, and have to reflect very carefully on the phenomenology to discover this, as I'll emphasize more later. This linguistic phenomenology tracks a difference in deeper phenomenology. There is a feeling of *externality* to the kind of striking reported by some uses of (1): that P simply presents itself, as if from the outside. No investment in P's truth is thereby attributable to one, and the reason seems to be precisely this special sense of externality. Hence the use of "strikes", and the distinctive syntax: "It strikes me as true" inverts "I am inclined to think it the case that P".

This is like perceptual experience, as some recent writers have noted.116 Perceptual experience also has a *presentational* phenomenology. This is a big part of what distinguishes it from mere imagination. The presentationality of perceptual experience doesn't *merely* consist in the fact that the contents of such experiences are seen *as real*. That also holds for belief. The relevant presentationality partly consists, as John Bengson notes, in their external relation to the agent:

> Having a presentational state such as perceptual experience […] can thus be understood as a *happening*, i.e., something that *happens to* one. Contrast imagining, guessing, hypothesizing, believing, and judging, which typically do not *happen to* one […]. The implication is not that these other states (non-happenings) are wholly voluntary while only presentational states are non-voluntary; rather, the point is that there are varieties of non-voluntariness, and presentational states such as experiences […] are non-voluntary in ways that beliefs, imaginings, guesses, hypotheses, or judgments, are not.117

115See Koksvik (ms) for arguments like this. I had the thought before discovering his far more detailed work. Anyone who wants to think about these issues should read him. He has a whole dissertation on the matter.

116See Chudnoff (2011) and Bengson (ms).

117Bengson (ms: 28).

This goes hand in hand with our first observation—viz., that one cannot be rational or irrational for having a perceptual experience, or having an intuition of the kind indicated by (1). Indeed, a natural thought is that the distinctive presentational character of these states *explains why* one can't be rational or irrational for undergoing them, though one may be "defective" in other ways.

A third difference between some intuitions and beliefs, credences, inclinations to believe, and attractions to belief lies in their *fineness of grain*. Bengson sets us off in the right direction here:

> [Consider] a "novice" possessing native mathematical abilities (comparable to Ramanujan's), who prior to receiving formal training shows early signs of genius; she might have an informal mathematical insight about how to solve a particularly difficult proof, but she lacks the formal tools required to formulate (even to herself, "in her head") what she saw. In such a case, the content of her intuition may be highly specific or determinate, e.g., *that the way to solve the proof is by …*; but the content of her attraction—that to which she is attracted to assenting—may be something less specific, e.g., *that there is a way to solve the proof, and it has to do with….* Thus she might have to poke around for a bit until she hits on the solution (or its exact articulation), at which point she might exclaim, "That's it; that's what I saw!" Such an exclamation would provide some reason to think that the content of her intuition—what she saw—was in fact highly specific, even while the content of her attraction was not.[118]

Although Bengson's case is unusual, insight of this type is, I believe, hardly rare. We see and exhibit it in having everyday normative intuitions of many different types.[119]

Here is one kind of example. Often very fine-grained features of acts strike us as making them wrong. Almost equally often, it is terribly hard to isolate and describe the reason-giving features that are striking us. So we resort to coarse-grained normative talk ("It's just wrong!"), even though remains clear that the facts picked out by this talk reduce to facts about a balance of some specific reasons. All the same, one is *de re* aware of the reasons, and is struck *by them*. One is even in a position to respond to them in many cases where they remain difficult to isolate. This is brought out by the fact that one can have very different reactions to slight variations. Thus the familiar methodology in ethics: consider alternative cases that differ only in minute key details from the paradigms, which hopefully evince what struck us from the start. We manifest our sensitivity to certain kinds of reasons "offline" in being struck as we are. Subsequent theorizing is devoted to figuring out what the specific reasons that struck us (*de re*) really were.

This resembles the attempt to recapture in a painting or long poem or story all the details that are so hard to pin down in our perceptual experiences. This familiar point leads to the observation that perceptual experiences are rich in content in ways that go beyond what we can get into our beliefs on the spot. The precise details of this content can strike us in certain ways, even if we are unable to put our finger on the details doing the work. Our reactions to cases in ethics and in other walks of normative reality are rich in the same way. Since our beliefs, credences, inclinations to believe, attractions, and so on, are not directed at propositions with a remotely similar degree of specificity, the two are distinct. *Pace* Sosa (2007, 2009), their contents can diverge significantly. Indeed, this is to be expected on other grounds, given the amount of disagreement there can be in philosophy about the content of certain intuitions. Often the hardest task is to figure out

---

[118]Bengson (ms: 49).

[119]I'll focus just on moral examples for vividness. I think there are examples structurally like this across the board. I use them to deflate quick arguments for "buck-stopping" accounts of the relation between values and reasons in Sylvan (ms3). Some good non-moral examples are aesthetic ones. In hearing Bach's *Chaconne* for the first time, one is rightly awestruck. One senses many reasons to feel like this toward this sublime work, even if one is bad at isolating and describing them. One can bring out how particularized one's sensitivity is by considering variations on the music that differ in slight ways. For these reasons, the variations may not correctly inspire awe in one, or as much awe as the *Chaconne* does.

precisely what the intuition really is supposed to be. Subtle but important mistakes are made when the content of the intuition is not correctly formulated.

These are three main differences that lead me to deny that one salient type of intuition can simply be identified with belief, credence, or inclination to believe. One can be rationally criticizable for all of these latter states. None of these latter states are presentational in the relevant sense. Very often, none of these latter states can have the same kind of rich content. These are the kinds of states I have in mind in discussing nondoxastic appearances. Given the important differences, I can't see that the defender of the Weak Appearance-Relative View faces the objection to the Weak Doxastic View, at least when that objection is properly understood.

Many normative intuitions are fine examples of these kinds of states, even if their intuitionist friends in ethics have not understood them well. They often figure in capacities we exercise routinely in ordinary life, online and offline. A revealing case is afforded by emotion. For many terms that pick out emotive attitudes—'shame', 'disgust', 'contempt', 'amusement', 'awe'—there are corresponding normative terms, like 'shameful', 'disgusting', 'contemptible', 'amusing', and 'awesome'. To be awestruck plausibly just is to have a feeling of with certain character struck up *by a spontaneous, fine-grained presentational representation of an object as having features that merit this attitude (awe).* 120 Some normative intuitions *just are* the kinds of spontaneous presentational representations that can *partly constitute* these emotions, triggered offline. We can know that facts or things have the corresponding normative features by tuning our beliefs to the presentational representations that are bases for the feelings, which are the other constituents of emotions.121

There is nothing mysterious about the epistemology here. It is a familiar part of everyday life. If there is mystery, it is in the nature of the represented reason-giving relations. The epistemology itself is as unmysterious as any reasonably accurate normative epistemology could be. Ours is certainly no naïve, "bogus intuitionism" (to quote McDowell) that models normative epistemology on perception of primary qualities, as G. E. Moore perhaps had it. The problem with "bogus intuitionism" is not its stress on *intuitions* or the *perceptual analogy*, but rather with how it frames the analogy and understands the intuitions—as well as the picture of normativity it rests on, often one where normative features are seen as simple, primary, nonnatural qualities.

3.   *On to epistemic rationality: flawed antecedents and renewed prospects for a unified theory*

So much for practical rationality, and its distinctness from responsiveness to normative reasons, including possessed ones. What I will now turn to do is show how an exactly parallel distinction can be drawn between what we would be rational in believing and what we have justification to believe. There is a general account of rationality in the offing, as well as a general conceptual distinction that crosscuts domains. The point of advancing the literature on practical rationality and extending Reasons Basicness in Chapter 1 to epistemic justification was to set this up.

Before I turn to this, one might reasonably ask why it is not nearly as common for epistemologists to distinguish between rationality and justification as it is for philosophers of practical reason to

---

120A nonnormative analogy is found in the perception of secondary qualities. Seeing red plausibly just is having a veridical experience *that represents its object as causing experiences of this sort (phenomenally red).* Correctly emoting has a similar structure, though 'causing' needs to be replaced by 'meriting', as I've suggested. See "Values and Secondary Qualities" and "Projection and Truth in Ethics" in McDowell (1998) for this story. Cf. De Sousa (1990) and Jacobson (2005) for more detailed stories of this appealing type.

121Emotions have, I think, both a sensory side and a representational side. Maybe the sides are identical. On one kind of representationalism, they are. I assume otherwise because it seems to carve the joints better.

do so.122 This is a remarkable fact. Of course, I have my own big picture answer. I think there is a *substantive* disunity between the practical and the epistemic domains that guarantees that the *extensions* of 'rational' and 'justified' are closer in the epistemic domain, even though there ought to be a widely recognized and important *notional* distinction between them. Since our intuitions track this fact, it is easier to conflate these properties in epistemology. But it will take the entire dissertation to see this. This would be a theory-driven explanation.

Even without it, I think there are some clear reasons why epistemologists avoid the distinction. One reason is that the only prominent attempt to draw something like it in epistemology was flawed. Foley (1987, 1993) was interested in a project not unlike mine. He wanted to separate epistemic rationality from whatever normative standing is required for knowledge. His execution failed, and many epistemologists found it unacceptable. 123   The fact that his specific implementation failed may have made other epistemologists skeptical about the more general idea, or about the usefulness of the ordinary senses of "rational" and "irrational".

It is worth seeing how his view was flawed, so that we can avoid his mistakes. Like me, Foley started by thinking in general terms about rationality, and then tried to construct an account of epistemic rationality that is subsumed by the general picture. One of his mistakes was his starting point. His "Aristotelian" view of practical rationality was *teleological* and *coherentist*:

> By an Aristotelian conception of rationality, I mean one that is goal-oriented, one that understands rationality in terms of a person carefully deliberating about how to pursue his goals effectively and then acting accordingly. […] [L]et us broaden it by thinking of rationality in terms of what a person would have to do in order to pursue his goals *in a way that he would believe to be effective, were he to be carefully reflective* […]. Thus, according to the Aristotelian conception, rationality is best understood in terms of a person pursuing his goals *in a way that he would believe to be effective were he to take time to reflect carefully* […].[124]

> Does this way of thinking about rationality suggest that the goals of an individual are not themselves susceptible to rational evaluation? […] No; given his other goals, it might be very irrational for him to have X as a goal, since X might interfere systematically with his other goals. *What this way of thinking about rationality does preclude is the possibility of an individual's goals being coherent, in the sense that there are no systematic conflicts among them, and at the same time being thoroughly irrational. So, if we imagine an individual with coherent albeit perverted goals, this way of thinking about rationality may not allow us to say that it is irrational for the individual to have these goals.*[125]

Foley takes a *reflective* version of the *Subjective Instrumental Principle* to be the core of rationality:

> **SUBJ-R**: Rationality requires that if S has total goals $\{G_i\}$, and would after careful reflection regard A-ing as an effective means to some goal(s) in $\{G_i\}$ and not a significant hindrance to any other goals in $\{G_i\}$, then S pursues A-ing.

122The clearest exception is Jackson (2011), who is well aware of the new literature on rationality, and who puts his knowledge to great use in criticizing a certain kind of argument for "seemings internalism" that he calls the "Argument from Irrationality". I shall later discuss some of Jackson's remarks about inferential rationality, since it is the one place—by my lights at least—where there is an initially more plausible disconnect between responding to apparent reasons and epistemic rationality in the narrow sense.

123See the September 1989 issue of *Philosophy and Phenomenological Research* (Volume 50, No.1) for a symposium on Foley (1987). The contributions to the symposium by Alston, Feldman and Swain all converge on similar objections to the ones I'll be mentioning below. Virtually everyone in the literature I've seen discuss Foley's theory at any length make similar objections.

124 Foley (1987: 5−6); italics mine.

125 Foley (1987: 7); italics mine.

One might ask what "careful reflection" involves in SUBJ-R. "Careful" sounds normative. So perhaps there are requirements that spell out what it takes to reflect carefully. Strikingly, however, this is not Foley's view. It takes a while to see this: he tells us nothing about carefulness until he states his account of epistemic rationality. "Careful" and "sufficient" reflection, as we'll see, do not for him involve compliance with further requirements: they are cashed out belief-relatively. Foley's implicit view of rationality is *purely coherentist*: it requires means-end coherence, relative to beliefs about means-end relations with which one would be *satisfied* on reflection.

Now *epistemic* rationality for Foley is restricted practical rationality, where the goal G is just the epistemic goal of *now having true beliefs* and *not now having false beliefs*. He seems to add further constraints, but these are supposed to flow from the content of the goal. Here is the core idea:

> [I]t is epistemically rational for an individual to be persuaded of the truth of just those propositions that are conclusions of arguments that he would regard as likely to be truth-preserving were he to be reflective and that in addition have premises that he would uncover no good reason to be suspicious of were he to be reflective. The idea, in other words, is that a proposition is epistemically rational for a person just if using premises that are uncontroversial for him, he can argue for the proposition in a way that is uncontroversial for him.[126]

Two crucial elements of his account are (i) that, in non-foundational cases, a subject can rationally believe P only if he would on careful reflection regard the arguments on the basis of which he would believe P to be *sufficiently likely to be truth-preserving*, and (ii) that, in these cases, a subject can rationally believe P only if he on careful reflection would uncover "no reason to be suspicious" of the premises of the argument for P. Foley doesn't think arguments go all the way down. Eventually we hit foundational premises that the person "uncovers no good reason to be suspicious of". This is why he calls his view a kind of "subjective foundationalism".

Now, "no good reason" here implicitly means, for Foley, "no reason that the subject would regard as a good reason after sufficient reflection". This is clear when we turn to his account of what it takes for a subject to engage in *careful reflection*:

> But in what kind of careful reflection are we to imagine a person engaging? The answer is: Reflection that reveals *the person's own deepest epistemic standards*. This answer, of course, merely pushes the question back a step. What kind of reflection reveals a person's own deepest epistemic standards? The answer is: Sufficient reflection from an epistemic point of view.[127]

Foley goes on to cash out "sufficient reflection from an epistemic point of view" as follows:

> How much reflection is required before his opinion of an argument is indicative of his own deepest epistemic standards? Is it enough for him to reflect for a few minutes, or for a few hours, or for a few days, or what?
> The answer is that strictly speaking there is no limit. We imagine him reflecting *until his view stabilizes, until further reflection would not alter his opinion of the argument in question*. This by definition is the point at which the person is not susceptible to further self-criticism. It is by definition the point at which had the person reflected still more, he would not have charged himself with being mistaken in his earlier evaluation of the argument.[128]

Here we see at last that "sufficient reflection" and "careful reflection" are just sufficient and careful reflection *by the subject's own lights* when the subject cares only about epistemic goals. So there are indeed no further requirements of rationality associated with carefulness for Foley.

---

[126] Foley (1987: 5).
[127] Foley (1987: 33).
[128] Foley (1987: 35); italics mine.

The core problem is that this creates too much subjectivity, even if we are talking about rationality in the ordinary, narrow sense. Just imagine someone whose beliefs would stabilize after virtually no thought, or rigidly dogmatic people who disregard apparent reasons they don't antecedently acknowledge. Nothing so far blocks such people from being perfectly immune to rational criticism on Foley's view. One can imagine a person of this sort staring contrary appearances in the face without the slightest alteration in his beliefs. Even in the ordinary narrow sense, surely such a person could be irrational. This casts doubt on Foley's kind of view.

Perhaps Foley cannot be blamed. Excessively subjective accounts of practical rationality had been the norm. Foley's SUBJ-R is a lot like the deliberative subjectivism that originated in Brandt (1979) and Darwall (1983), and that many others have since adopted. This is why I wanted to put forward my own account of practical rationality before trying to secure my analogies between the practical and epistemic spheres. Part of what hinders progress in importing the relevant distinctions into epistemology is that some very common ways of drawing them in the practical literature are defective. When one applies the unmodified analogies, one gets an intolerably subjective view. Precisely because of this, many epistemologists may have found the ordinary sense of 'rational' and 'irrational' uninteresting. Foley does not help to discourage this. Crucially, though, this is no strike against the general ambition of articulating the difference between the ordinary narrow notion of rationality and justification, or of seeing core structural similarities between the epistemic and practical spheres.

What else might be responsible for the failure to draw a distinction between epistemic rationality and a less narrow normative standing, like having sufficient epistemic reasons? I am tempted to think that there is just a kind of dialectical inertia whose magnitude owes to sociological factors. Sometimes in philosophy, disputes are framed in the wrong terms from start. Because disputants state their views in these terms, and want a smooth ongoing discussion to which they can contribute, they decide not to change the terms of the disagreement. The longer they wait, the harder it becomes to change the terms this without seeming too "far out" or deviant.

Precisely this is true, I believe, of internalism/externalism disputes in epistemology. They started as disputes about epistemic justification. But one side started in confusion. I noted in the Introduction how many internalists simply slide between talk of justification and talk of rationality. They sometimes even claim this to be *definitional.* Consider Huemer:

> Another word for what is justified, or *should* be done from the first-person perspective, is "rational".[129]

Stewart Cohen was even clearer in framing his "new evil demon" problem for simple reliabilism:

> '[R]easonable' and 'rational' are virtual synonyms for 'justified' […]. If the Reliabilist wants to distinguish 'justified' from 'reasonable' or 'rational' he may do so. But clearly the important epistemic concept, the one epistemologists have been concerned with, is what the Reliabilist would call 'reasonability' or 'rationality'.[130]

The first claim is false, as we can see on general grounds. So is the last claim. What epistemologists were *primarily* interested in for many years was just the *normative standing that is required for knowledge.* Often "justification" was *stipulatively defined* by its theoretical role as one of the necessary conditions for knowledge. It is far from clear that once we understand epistemic rationality on its own terms, it will fit this role. It is also far from clear that if we understand epistemic justification on its own

---

[129] Huemer (2001: 22).
[130]Cohen (1984: 283−4)

terms, it will coincide with rationality. General considerations and analogies with the practical case suggest that exactly the reverse will be true in both cases.

Now, Foley was keen on examining epistemic rationality on its own terms, and avoiding the (stipulative) presupposition that it will have a role in the analysis of knowledge. Although I reject his theory, his keenness was fitting. He spoke with some insight in saying:

> There are two strains in the Cartesian project. One is that egocentric rationality is essentially a matter of being sufficiently reflective and cautious in thinking about what to believe. The second and very different strain is that if we are sufficiently reflective and cautious, we can be assured that we won't be led into error. The Cartesian project fails because we cannot have it both ways. There is no method that is at once subjectively convincing and objectively guaranteeing. These two aspects of the Cartesian project cannot cohabit in a single notion of rational belief. The question is how to react to this.
>
> One reaction is to abandon epistemology. […] But of course, this isn't the only option. […] [A]n obvious strategy is to split the Cartesian project into two. One project is to pursue the objective part of the Cartesian enterprise, the part that is predominantly concerned with the conditions of knowledge. Those who pursue this project will be asking how we as inquirers must be related to our environment if we are to have knowledge. […] [H]owever we answer this question, whether it be in terms of using reliable methods or in terms of our cognitive equipment's functioning properly or in terms of our having adequate objective evidence, we will not have given a satisfactory answer to the questions that arise out of the egocentric predicament. […] No account of knowledge can capture this egocentric notion of reasona[bility] and the associated egocentric notion of rationality […].[131]

Foley is right that there is a tension. It emerges clearly when we reflect generally on the difference between justification and rationality, which card-carrying members of the Cartesian tradition have conflated time and time again. And Foley didn't take the idea far enough. In his (2004), he went on to characterize what was needed as a trial separation of the *theory of justified belief* and the *theory of knowledge*. He followed others in conflating justification with rationality.

Foley's earlier thought did not catch on among other internalists. Both conflations persist. I suspect part of the reason is just inertia and the desire for there to be a common terminological ground that guarantees at least the semblance of a disagreement. Consider Fumerton:

> [T]he metaepistemological project I am interested in concerns the concept of justified or rational belief. I have acknowledged that the expression "rational" might be somewhat less misleading than the expression "justified," but I will continue to use the two terms interchangeably in this work, in part because by now there are a number of technical expressions using variants on the word "justified" that are an almost indispensable part of the literature on such controversies as foundationalism.[132]

This terminological inertia and the conflation on which it rests has to change if we are to make progress. In the epigraph at the beginning of this chapter, Judy Thomson was talking about terms that seem *even more closely connected* than 'justification' and 'rationality'—'ought', 'must', 'correct' and 'obligation'. If it was, as she says, "just a mistake" to mix these up into a "nice warm conceptual pudding", it is a greater mistake to do the same with 'rational' and 'justified'.

As I emphasized in the Introduction, this does not mean that there are no interesting debates to be had between the familiar line-up of internalists and externalists—Foley, Fumerton, BonJour, Huemer, Pryor, and some of the responsibilist virtue epistemologists on one front, and Goldman,

---

[131] Foley (1993: 85).
[132] Fumerton (1995: 19).

Kornblith, Bergmann, Plantinga, and some of the reliabilist virtue epistemologists on the other. The debates simply have to take a different shape. Exactly what shape they should take will become apparent in the next chapter, when I discuss the problems of deontic significance for practical rationality, and generalize these problems for epistemic rationality. The interesting questions are about *how* rationality could bear on justification. They are about whether believing irrationally does anything *per se* to contribute to believing unjustifiedly, and whether believing rationally does anything *per se* to contribute to believing justifiedly.

These are hard questions. What people with internalist sympathies ought to do is find a fundamental epistemic norm that explains how rationality could be truly deontically significant, rather than a *merely* hypological property. As I'll argue, what needs to be undermined to show rationality to have *any* deontic significance is the tradition of viewing the fundamental epistemic norm as the teleological one of truth-conducivity and falsehood-obstructivity. Unless this goes, radical externalism holds for the deontic in epistemology—at least barring radical skepticism.

In the pages that follow, I will start to break the inertia. This may sound like a hopeless task from a sociological point of view. Perhaps it is. But we must remember that a similar inertia plagued the practical literature until the late 1990s. The shift out of it has taken place very swiftly. As in epistemology, the internalists were largely to blame.[133] Ironically, the same Bernard Williams who gave us the kind of case that seems most helpful in prying apart reasons and rationality was himself a primary culprit. Williams wrote:

> It is the external reasons theorist who faces a problem at this point. There are of course many things that a speaker may say to one who is not disposed to φ when the speaker thinks that he should be, as that he is inconsiderate, or cruel, or selfish, or imprudent; or that things, and he, would be a lot nicer if he were so motivated. Any of these can be sensible things to say. But one who makes a great deal out of putting the criticism in the form of an external reason statement seems concerned to say that what is particularly wrong with the agent is that he is *irrational*. It is this theorist who particularly needs to make this charge precise: in particular, because he wants any rational agent, as such, to acknowledge the requirement to do the thing in question.[134]

People have lately followed Scanlon in thinking that "Williams is quite right that this claim would be implausible, but wrong […] to hold that his opponent is committed to it."[135] Wrong, as Scanlon rightly insists, precisely because claiming that there is a reason for someone not to do something needn't imply that this person would be *irrational* in doing it. Rationality and irrationality are narrow and perspectival in a way in which facts about what reasons there are for people to do things are not. If this change can happen so quickly, and do so much good in the philosophy of practical reason, one can hope that the same thing might happen in epistemology.

3.1.   *The weak appearance-relative view extended*

It is easy to see how it could be done in virtually the same terms, at least at one level, now that we've corrected our understanding of practical rationality. Epistemic rationality, like practical rationality, partly consists in responsiveness to *apparent reasons*, where "apparent" means "relative to ultimate nondoxastic appearances *vis-à-vis* what reasons there are". The apparent reasons are

[133]Some were the same people! Fumerton (1990)'s arguments for a radically egocentric theory of practical reasons and for skepticism about commonsense morality rested in an obviously essential way on an explicit conflation of A-ing rationally and A-ing in line with the reasons there are or that are possessed. It is remarkable to look back on works like this and see how fundamentally the terrain has changed.

[134]Williams (1981: 110). I take from Scanlon (1998: 27) the idea of using the quotation to illustrate this conflation. He points to several other internalists who made the same mistake in Scanlon (1998: 27–9).

[135]Scanlon (1998: 27).

simply epistemic rather than practical. As with the practical case, there will be aspects of epistemic rationality that we cannot capture just with a requirement to respond to apparent reasons. WARea**, WAR1ep and WAR2ep were not the whole story about practical rationality. We also needed a version of the Subjective Instrumental Principle and a host of wide-scope requirements of coherence like Intention Consistency and Enkrasia. We'll end up needing a similar handful of further requirements to fully limn epistemic rationality.

How should we extend the Weak Appearance-Relative View? For *ex ante* epistemic rationality, we could just try to make the relevant substitutions in WARea**, and uphold:

> **WARe**: Assuming satisfaction of all wide-scope requirements of coherence for belief,
>> S would be *ex ante* epistemically rational in believing that P iff
>>> (i)   according to the total ultimate nondoxastic appearance for S, there are (at least) sufficient epistemic reasons for believing that P, and
>>> (ii)   S is in a position to believe P for these apparent epistemic reasons in the secondary strong sense,
>
> and
>> S would be at least somewhat *ex ante* epistemically irrational in believing that P iff
>>> (iii)   according to the total ultimate nondoxastic appearance for S, there are (at least) sufficient epistemic reasons against believing P, and
>>> (iv)   S is in a position to disbelieve P or suspend judgment on P for these apparent reasons in the secondary strong sense,
>
> where degrees of (ir)rationality are proportional to the strength of the ultimate appearance.

But we can actually be a bit more informative than this, and simplify these formulations.

To see how, start with the familiar fact that doxastic deliberation is *transparent*: deliberation about whether to believe that P collapses into deliberation about whether P.[136] An epistemic subject will see as reasons for a belief only factors bearing on the truth of its content, and will also see factors bearing on the truth of P as candidate reasons for or against believing P. This suggests that we can switch from talk about *what epistemic reasons for belief the ultimate nondoxastic appearances represent* to simpler talk about what facts the ultimate nondoxastic appearances represent to be *true* or *likely to be true*, *conditional* on the contents of these appearances. So we can go on to say:

> **WARe***: Assuming satisfaction of all wide-scope requirements of coherence for belief,
>> S would be *ex ante* epistemically rational in believing that P iff
>>> (i)   it is sufficiently likely that P conditional on the content of the total ultimate nondoxastic appearance for S, and
>>> (ii)   S is in a position to believe P for the apparent reasons in the
>
> make (i)   total content of the ultimate nondoxastic appearance that
> sense.   true, and to believe for these in the secondary strong
> and
>> S would be at least somewhat *ex ante* epistemically irrational in believing that P iff
>>> (iii)   it is not sufficiently likely that P, conditional on the content of the total ultimate nondoxastic appearance for S, and
>>> (iv)   S is in a position to disbelieve P or suspend judgment on P for the apparent reasons in the total content of the ultimate nondoxastic appearance for S   that make (iii) true, and in the secondary

---

136 Cf. Shah (2003, 2005, 2006) for the canonical discussion of this fact and its importance.

strong sense.

> where degrees of (ir)rationality are proportional to the strength of the ultimate appearance.

Exactly what "sufficiently likely" means is, of course, up for independent substantive dispute on all sides. I will simply be neutral on this background question, which is a question for everyone.

When we couple this picture with the account of normative epistemic reasons I defended in the first chapter, we see many parallels with the practical case. Epistemic reasons are *facts* in the good case, and *mere propositions* in the bad case. Where do appearances enter into the explanatory order? In the *good* case, *veridical* appearances function as *background conditions* which partly explain why certain facts are good reasons for someone to believe that P. In the basic case, it is in virtue of the ultimate appearance's veridically representing that P; in the nonbasic case, it is in virtue of the ultimate appearance's overall representational content making it likely that P.

To be a truly good reason, a content plausibly has to be *true*. This ought, at any rate, to be the default view, though I will ultimately dispute it. Otherwise the agent has, on the face of it, *merely apparent reasons*. The demon-deceived subject who presupposes that the world is as it seems is in a position to believe that there is a cup in front of him for the reason that there is a cup in front of him. It's just that he would be sadly mistaken about that, to echo Dancy. This content can be his reason all the same. Moreover, he would be rational in believing for it—though not on the face of it *justified*, since he would plausibly fail to believe for a *good* reason. This is exactly like the practical story. The conceptual gap between good reasons and rationality is exactly parallel. With argument, one can *try* to claim that the appearance itself functions as a background condition on a *sufficiently good* reason. But this is *far* from trivial, given the practical analogy.

Notice that I've duplicated the earlier strategy of assuming *ex post* rationality to be the more fundamental idea by adding clauses (ii) and (iv). This is good. It helps us avoid obvious concerns that crop up for any view that makes the contents of nondoxastic appearance play a big role. One concern is the "problem of the speckled hen". Consider a case that illustrates it:
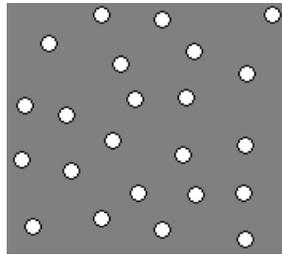
SPECKLES

Upon *simply viewing* SPECKLES for the first time, one can plausibly be rational in believing that there are *some* speckles. But one cannot plausibly be rational in believing that there are 23 speckles. Not without counting. Simply seeing SPECKLES does not make it *ex ante* rational. Yet one's experience represents both facts. So, on the face of it, both propositions ultimately nondoxastically appear true, and both are eligible apparent reasons. If we want to defend the Weak Appearance-Relative View, we can consider the following question:

> **Q**.   Granting for the sake of argument that a visual experience of SPECKLES gives you an apparent reason to think that there are 23 speckles before you, could you *believe for*

this apparent reason in the relevant sense (i.e., the secondary strong sense)?

A defender of my view can sensibly say 'no' to Q. Remember that in order to believe that P for some apparent reason in the secondary strong sense, one must be in a position to track what makes it the case that there is an apparently good reason. The appearance *of a good reason* must be explaining why one believes as one does. What makes having an experience as of SPECKLES give an apparently good reason for believing that there are 23 speckles is that *it represents 23 speckles*. But one cannot track *this* fact in the required sense just upon having the experience. This is *guaranteed by one's limited discriminatory abilities*. So what makes the experience rationalize believing this content can't relevantly explain why one believes there are 23 speckles. The defender of our view can insist on principled grounds that one here can't be *ex post* rational in believing this proposition *for the apparent reason given by the experience* simply upon having the experience. He can also claim that one could not be *ex ante* rational, given that *ex ante* rationality amounts to being in a position to believe with *ex post* rationality, given clause (ii) in WARe*.

Notice that this is exactly parallel to the way in which we avoided the "problem of difficulty" for the Weak Appearance-Relative View about practical rationality. The problems are structurally exactly the same. To avoid the problem for our view about practical rationality, we appealed to the deep kind of act-agent blur that infects evaluations of rationality to motivate taking the *ex post* sense to be primary, and we noted that one could not be *ex post* rational in the problematic cases for principled reasons—reasons that emerged a long time ago when we discussed the strong sense of A-ing for a good reason in Chapter 1. We have all the same reasons for taking *ex post* rationality to be primary in the epistemic domain. Evaluations of epistemic rationality are like evaluations of practical rationality in exhibiting a deep kind of act-agent blur—one that probably owes, as I hypothesized at the outset, to the fact that any act-oriented rationality evaluations must be parasitic on underlying agent-oriented rationality evaluations.

3.2. *Dimensions of epistemic rationality beyond "saving the appearances"*

So far, then, we have a unified account of epistemic and practical rationality. Both are narrow and perspectival, and the way to explain these features is by viewing both as being keyed to *apparent reasons* in a restricted nondoxastic sense of "apparent". This move is recommended in part by our objections to Pure Doxastic Views, and in part by the fact that any evidence-relative view would have to rest on a conflation of responsiveness to genuine reasons and rationality.

But as with practical rationality, there are aspects of epistemic rationality that go beyond responsiveness to apparent reasons. Let me proceed to consider some of these.

3.2.1. *The dimension of intelligibility: wide-scope requirements and consistency*

One of these aspects is exactly parallel to a dimension of practical rationality discussed in §2.3. Practical rationality includes a dimension of bare intelligibility. The requirements of intelligibility were wide-scope requirements of coherence. Epistemic rationality includes such an element too, and these requirements are structurally analogous to the practical ones. In addition to practical enkratic rationality, there is epistemic enkratic rationality, embodied in wide-scope norms like:

> **Epistemic Enkrasia-WS 1**:   Rationality requires that one not simultaneously believe that there are decisive epistemic reasons to believe P and fail to believe P.

> **Epistemic Enkrasia-WS 2**:   Rationality requires that one not simultaneously believe that there are sufficient reasons to disbelieve that P and believe P.

There are many requirements like this. They reflect Scanlon's observation that a clear case of irrationality is failure of alignment between one's judgments about reasons and one's attitudes. What further wide-scope norms are plausible requirements of epistemic rationality? One *might* think that a requirement banning logical inconsistency will figure among the requirements. But we must be extremely careful in endorsing this idea. For it simply isn't true that all inconsistency makes for irrationality in the ordinary narrow sense, and it is only irrationality in this sense that requirements of rationality ban. Here is an obvious kind of example. You believe that P. A proof of untold complexity is required to see that P entails ~Q. You also happen to believe Q. Are *you* irrational? When we hear this as we ought—as an agent-oriented question, one about *your* openness to criticism—I think it is extremely implausible to say "yes".

One might say that the requirement will involve some kind of appeal to *clear* inconsistency. But what one might reasonably wonder is whether it must necessarily involve *real* inconsistency. Take someone encountering a paradoxical set of propositions—say, someone just starting to read Davidson's "How is Weakness of the Will Possible?" After having defined "incontinent A-ing" as *intentional A-ing while believing that there is an open alternative B that would be all-things-considered better to do than A*, Davidson pointed to the apparent inconsistency of these three plausible claims:

1.   If S wants to A more than B and he believes himself free to do either A or B, then S will intentionally A if he does either A or B intentionally.
2.   If S believes that it would be better to A than B, then he wants to A more than to B.
3.   There are incontinent actions.

These claims really do look inconsistent, if you understand Davidson's definition and aren't as ingenious as he is. Upon realizing this, it is plausible that one is rationally required to suspend judgment on the conjunction. Nevertheless, assuming that Davidson was right, the conjunction does not really entail a contradiction. So suppose Davidson was right. Then one can be rationally required to respond to merely apparent inconsistency as if it were real inconsistency.

Real inconsistency seems to be neither necessary nor sufficient for irrationality. This suggests that there are not any pure requirements of consistency for the ordinary notion of rationality. Indeed, we don't even need to go beyond the requirement to respond to apparent reasons to capture what is going on in the cases where (apparent) inconsistency does seem irrational. When it appears that there is inconsistency, it *ipso facto* appears that there is definitely some falsehood. And its appearing that *if* one believed all three propositions one *would* definitely believe something false is enough to rationally require one not to believe all three propositions, simply because it would appear that there is a conclusive epistemic reason against the collective endorsement. So if we're interested in finding requirements of epistemic rationality beyond the requirement to respond to apparent reasons and to be "enkratic", we must look elsewhere.

3.2.2.   *Inference and a puzzle about the rationality of deductive* versus *non-deductive cases*

One place to look is at evaluations of belief formed *via* basic deductive inferences. Here there arises a fascinating apparent tension in our thinking about epistemic rationality. Rationality gets intuitively more demanding here. There is a task in explaining this consistently with earlier ideas. Of course, there are narrow requirements on deductive inferential rationality, like Broome's

> **Modus Ponens Requirement**: Rationality requires that: if S believes that P, and believes that if P then Q, and cares about whether Q, then S believes that Q.[137]

---

[137]Broome (Ms: 159).

Being wide-scope, there are many ways to satisfy this. One can drop belief in P, drop belief in the conditional, stop caring about whether Q, or go on to believe Q. It is in this respect a weak requirement, and belongs with incoherence-banning requirements like the enkratic requirements. Predictably, I think there is a stronger wide-scope principle than Broome's—viz., the

> **Appearance-Relative *Modus Ponens* Requirement**: If it ultimately nondoxastically appears to S that P entails Q, then rationality requires that if S believes that P, and cares about whether Q, S believes that Q.

But what is fascinating is that basic deductive inferential rationality appears to require more. For some subjects, even the ultimate appearance *vis-à-vis* the truth of a simple conditional spelling out a principle of deductive reasoning could mislead. Yet as Alex Jackson notes in an insightful discussion that acknowledges the gap between justification and rationality, "[t]he product of a fallacious inference is not normally called 'rational'."[138] He is using the familiar narrow sense of 'rational' here. What he says rings true with me even when I self-consciously restrict my attention to the relevantly narrow sense. Suppose the conditional at issue is one spelling out a fallacious inference. *Even if* the subject—say, an undergrad Tortoise—isn't in a position to see the fallacy on reflection, the subject would seem irrational in inferring according to the rule.

Rationality seems more closely bound up with correctness here. One can *try* to insist that we are making another conflation of rationality with a more substantive idea. It would be nice if that worked. But the obvious disanalogy with ampliative inference strikes me as requiring one to say a bit more. If enumerative induction turns out to be radically unreliable, that fact by itself is irrelevant to whether our inductive inferences are rational. All we need is an *ultima facie* appearance of reliability. But an appearance of validity in the basic deductive case is *not enough*, even if the appearance is ultimate and strong. Indeed, an appearance of significant reliability that could in *rare* cases go wrong isn't enough, if one is really using the rule *in deduction*.

What is going on here? It is worth remembering that good ampliative rules are formally invalid. When I first learned about inference to the best explanation in a philosophical context in my intellectual infancy, I remember thinking—steeped as I was in freshman deductivism—"Isn't that clearly no good? It's affirming the consequent!" The silliness of this now seems easy enough to see. The inference was never presumed deductively valid. Its quality surely *wittingly* depends on contingent facts about the world as much as the quality of enumerative induction. Something within the explanation of why the thought was foolish is helpful. Clearly, if one were engaging in affirming the consequent with a witting presumption that one is engaging in an ampliative rather than a deductive inference, one could not be faulted for irrationality *simply* for using a deductively invalid rule. So Jackson's observation needs to be restricted: only fallacious reasoning *with a presumption of deductive validity* can open one up to a charge of irrationality.

Why not, then, simply say that as long as the presumption is itself rational, one will be free from the charge of irrationality? The trick is to explain why the standards governing the presumption seem more demanding than is otherwise typical of rationality in the familiar sense. Part of the interest of Jackson's point is that *even when* it appears to one—ultimately appears, let's say, simply because one is very dull—that a deductively invalid inference is valid, one is not fully rational in going ahead and deploying the inference. Why are the standards for avoiding irrationality higher? They are clearly not this high for wittingly non-deductive inferences. It cannot simply be that some of us could, by *a priori* reflection, distinguish the correct simple deductive rules as such. Not all of us could. Even those of us who cannot still seem criticizably irrational.

---

[138]Jackson (2011: 575).

I think the best solution to this puzzle invokes a modest kind of inferentialism about logical concept possession and identity. Fully understanding a logical concept plausibly involves accepting and being disposed to apply *in deductively presumptive inferential activity* the rules that constitutively govern it. When one engages in affirming the consequent *with a deductive presumption*, one is tacitly taking it for granted that the conditional is governed, as a matter of constitutive fact, by different rules than the actual constitutive rules. If there were no such presumption, we'd face a less demanding case that could be understood just as we understand the rationality of ampliative inference, which doesn't require more than responsiveness to apparent reasons or the wide-scope coherence demanded by the *Modus Ponens* Requirement.

Insofar as one continues to possess the logical concept, there will be a tension between the new presumption and the older acceptances in virtue of which one got the concept in the first place. The irrationality of a deductively invalid inference—wherever fairly placed—lies in this tension. It is a deep tension of just the sort that requirements of rationality are in the business of banning. Crucially, nothing like this holds for non-deductive inference.

Of course, there is a limit to the applicability of the charge of irrationality. At a certain point, after accepting too many new rules and failing to exercise or even possess a disposition to be governed by the constitutive rules for the relevant logical relations, one *lacks* the concept. At this point, one's defect is not a defect of rationality in the ordinary narrow sense. It is more substantive one—one of lacking the right concepts, or of presuming that there is a novel logical relation that could be individuated by what turn out not to be truth-preserving rules. If the appearance of truth-preservingness is strong enough in *that* case, I lose my grip on the idea that one would be truly irrational, rather than just suffering from a severe substantive defect.

Let me say a bit more about wherein the incoherence I claim to exist in our troubling case consists, and then more about how modest the inferentialism I'm invoking is intended to be. The constitutive rules for a connective are its Intro(duction) and Elim(ination) rules. For →:

$$
\begin{array}{c}
[A] \\
\cdot \\
\cdot \\
\cdot \\
B \\
\hline
A \rightarrow B
\end{array}
\quad (\rightarrow\text{INTRO}) \qquad\qquad
\begin{array}{c}
A \rightarrow B \qquad A \\
\hline
B
\end{array}
\quad (\rightarrow\text{ELIM})
$$

Like the Intro and Elim rules for all the familiar truth-functional connectives except possibly classical negation, these have a certain nice property. They are, to use Dummett's influential term, in *harmony*.[139] Informally, the Intro and Elim rules for a connective C are in harmony just when the consequences that could be drawn (in any case) with the Elim rules for C do not outstrip the grounds for introducing C (in any case) in the first place. As Stephen Read puts it: "If the elimination-rules do no more (and no less) than is justified by the introduction-rules, the rules may be said to be in harmony."[140] The rules that are constitutive of the conditional have this property: given →ELIM, the consequence of accepting 'A → B' is just that B can be inferred from A, but the very point that B can be inferred from A was what was needed to introduce 'A → B' in the

[139]See Dummett (1991) for the classic discussion. Dummett uses the idea to motivate accepting intuitionist logic precisely on the putative grounds that the rules for classical negation fail to be in harmony. People have, however, later disputed whether the properly articulated rules for classical negation must fail to be in harmony; see, e.g., Read (2000) and Gabbay (ms).

[140]Read (2010: 559).

first place, given →INTRO. There are various ways of trying to make this informal idea more precise—and they certainly matter if we want to avoid accepting intuitionistic logic!—but informality will serviceable enough for present purposes.

The idea of harmony allows us to make a bit more precise the idea I was gesturing at in saying that there is a kind of deeper incoherence in someone who applies affirming the consequent in deductively presumptive inferential activity.   In using this rule in deductively presumptive inferential activity, this person is in effect accepting a new elim rule for the conditional—namely:

$$A \rightarrow B \qquad\qquad B \qquad\qquad \text{(BAD} \rightarrow \text{ELIM)}$$

$$A$$

Unless this person *also* accepts a new Intro rule, his personal → concept will fail to have harmonious Intro and Elim rules. According to his Elim rule, a consequence of accepting A → B is that A can be inferred from B. This fails to fit with the Intro rule he must have accepted in order to acquire the concept of the conditional *ab initio*: the consequences that can be drawn with his Elim rules outstrip the grounds he had regarded as being the proper ones for introducing 'A → B' in acquiring the concept →. Of course, he could go on to accept a new Intro rule to ensure harmony—for instance, this rule:

$$[A] \qquad\qquad [B]$$
$$. \qquad\qquad\quad .$$
$$. \qquad\qquad\quad . \qquad\qquad \text{(MISLABELED BICOND)}$$
$$. \qquad\qquad\quad .$$
$$B \qquad\qquad\quad A$$

$$A \rightarrow B$$

But if he accepted this as constitutive of the concept →—something he would do if this rule were part of his deductively presumptive practice of using →—it is plausible that the concept → in his language of thought would just take on a new semantic value. It would pick out the *biconditional* rather than the conditional, which is defined by this further Intro rule, together with the Elim rules that he had already been accepting. Of course, I've continued to use the *symbol* '→' just to illustrate the origin of the shift in this person's language of thought.

So, there are two cases to consider. In one case, our subject has not yet accepted the new Intro rule. In the other, he has. In the latter case, he cannot be faulted with irrationality: he has switched to a different concept with which he *would* be reasoning correctly, though his beliefs may be false. In the former case, he can be faulted with irrationality *on grounds of disharmony*.

Why does rationality ban accepting disharmonious rules? By definition, when one reasons with an elim rule for a connective C that is disharmonious with the intro rules one already accepts for C, one reasons in such a way that the consequences one can draw (in any case) outstrip the justification one could have (in any case) for introducing a claim whose leading connective is C. The consequences outstrip the justification one had for introducing that claim just in case there is a possible case where one has sufficient reasons for introducing it without having sufficient reasons for some consequences drawn. But when one reasons *with a deductive presumption*, one reasons in a way that is presumed to be case-insensitively permissible, assuming that the premises of one's reasoning are themselves permissibly accepted. This presumption commits one assuming that having sufficient reasons for the premises would be enough, simply as a point of constitutive

necessity, for having sufficient reasons for the consequences drawn. There is clear incoherence between *that* presumption and one's acceptance of disharmonious rules, which by definition allow for possible cases where having sufficient reasons for introducing the connective would *not* guarantee having sufficient reasons for the consequences drawn in applying the elim rule. This is an incoherence one could avoid only by either rejecting the new Elim rule (in our imagined case), or by accepting an Intro rule that is in harmony with it. Notably, one would simply be *switching to a different concept* (e.g., of the biconditional) in the latter case.

Focusing on our example affords an illustration, and also clarifies the appropriate limits to the charge of irrationality. Suppose that one acquires the concept of the conditional →. In acquiring it, one must have accepted →INTRO as a governing principle for deductively presumptive inferential activity involving →. Accepting *that* amounts to assuming

> (X)   that one has sufficient reasons in a case to accept A → B if and only if one can   deduce B on the supposition that A in that case.

Suppose—as we must if our case is to be one where a charge of irrationality would be at all clear—that →INTRO is the *only* Intro rule one accepts. And suppose that one also affirms the consequent in deductively presumptive inferential activity. In doing this, one accepts BAD →ELIM as an Elim rule for → in deductively presumptive inferential activity involving →. Accepting that amounts to assuming *at least* that

> (Y)   that one can deduce A on the supposition that B in any case *if* one has
> sufficient reasons to accept that A → B in that case.

(X) and (Y) jointly imply that one can deduce B on the supposition that A in any case only if one can deduce A on the supposition that B in that case. On pain of incoherence, then, one is committed to accepting a deduction of B from A only if one accepts a deduction of A from B. This is a commitment that any subject would have sufficient ultimately apparent reasons to reject. Being able to deduce A from A&B hardly implies being able to deduce A&B from A! Any subject for whom this wouldn't be ultimately apparent would be a subject suffering from such massive substantive defects that evaluations of rationality and irrationality would be inapt.

There are two factors that explain the irrationality of affirming the consequent in deductively presumptive inferential activity. One of them is the requirement to respond to apparent reasons, which we invoked at the end. The other is a wide-scope requirement on *assumption*—viz.:

> (*)   Rationality requires that if one assumes that C iff D, and assumes that E if D
> (or: if D, E), one assumes that if C, E.

We applied (*) to get the result that assuming (X) and assuming (Y) commit one on pain of incoherence to assuming that one can deduce B on the supposition that A in any case only if one can deduce A on the supposition that B in that case. Notice that we are not assuming that rationality requires anything particularly strong in assuming the truth of (*). This requirement is like Broome's *Modus Ponens* Requirement. The difference is simply in the slightly greater logical complexity of the contents at issue, in the omission of the "you care" clause, and in the mental states to be related. The mental states are *assumptions* rather than beliefs. Assumptions are, I think, open to a slightly stronger requirement than beliefs: if you assume that A, and assume that A → B, rationality requires you to assume that B *whether or not* you care about whether B. Why? Simply because assumptions are in the first instance *tacit* rather than occurrent states, and so do not "take up intellectual space" in the way in which beliefs more plausibly do. The "you care" clause can thus drop out of the picture.

The result accords with our intuitions about how you can escape irrationality in this case. Being wide-scope, you can satisfy (*) in several ways in our case: (i) by dropping your assumption of (X), (ii) by dropping your assumption of (Y), or (iii) by going ahead and embracing the clearly crazy conclusion that you can deduce B on the supposition that A only if you can deduce A on the supposition that B. (i) would amount to rejecting the standard Intro rule for the conditional and replacing it with something else—say, the Intro rule for the biconditional. If you do this, you will cease to have the concept, and so will avoid irrationality simply by incurring a substantive defect or by switching to a different concept altogether. (ii) amounts to rejecting the bad Elim rule. This is the way of wisdom. (iii) would put you up against a different requirement of rationality— namely, the requirement to respond to apparent reasons. So, as long as we embrace a modest inferentialism about logical concept possession and identity, we can explain why rationality seems to be more demanding in the basic deductive case without making rationality depend on anything other than responsiveness to apparent reasons and requirements of coherence—(*) in this case, which is in the same family as the *Modus Ponens* Requirement.

How modest is the inferentialism I've invoked? Notice that I have *not* been engaging in the kind of project that Boghossian has explored in recent years of trying to explain how we are *justified* in using certain *deductively valid* inference rules (e.g., *modus ponens*) by appealing to an inferentialist account of concept possession and identity.[141] I have instead been interested in something considerably narrower: explaining why it is *irrational* in the narrow sense to use *deductively invalid* rules like affirming the consequent by invoking the same kind of apparatus. The success conditions for this project are weaker. As far as I can tell, this frees me from the concerns some have raise Boghossian's project—specifically concerns that *grant* the views about concept possession and identity for the sake of argument, and show that these are not enough to show that we are basically justified in inferring according to *modus ponens*.[142] No existing criticisms of which I'm aware could challenge the move that I've made from the views about concept possession and identity, together with my own views about rationality in the narrow sense, to the irrationality of inferring according deductively invalid rules.

The criticisms, then, would have to be of the account of logical concept possession and identity *per se*. My view has been modest. I have assumed only these claims:

> (i) For any logical connective concept C, C is partly individuated by its Intro and Elim rules.

> (ii) To possess a connective concept C by using some token concept *c*, one must accept the constitutive rules of C as applying in any deductively presumptive inferential activity involving *c* and aim to be governed by them in such activity.

(i) does not imply that *any* set of Intro and Elim rules picks out a connective concept type, and so is not open to the concerns that surround Prior's infamous TONK. This commitment of inferentialism is not very controversial. (ii) is the somewhat more controversial commitment. It's worth noting that even people who oppose inferentialism as a general view about concept possession have historically been fine with inferentialism about logical concept possession.[143] I'm

---

[141]See Boghossian (2008), and particularly "Blind Reasoning".

[142]See Schechter and Enoch (2006) for concerns of this kind.

[143]Fodor (1990: 110) was clear about this: "Since I think Kripke's objection fails […] I'm inclined to think that maybe there is *no* objection to the idea that '+', 'and', 'all', and the like have the meanings they do because they play a certain causal role in the mental lives of their users. This would, of course, be to accept a distinction in kind between logical and nonlogical vocabularies […]. Gilbert Harman somewhere suggests that to be a logical word *just is* to be a word of which a use-theory of meaning is true. That proposal strikes me as plausible." Fodor has since retracted this view because he thinks these accounts are *circular*; cf. Fodor (2004a,

in this camp: I think inferentialism is hopeless for other cases, but OK for the logical case.

I am aware of two main objectors to (ii). One of them is a recent time-slice of Fodor. As I note in n.143 below, I have already given an extensive reply to his objections elsewhere. Another objector is Williamson (2003). Williamson suggests that even great logicians can reject seemingly harmless constitutive rules for connective concepts. He give the example of McGee (1985), who thought he had discovered counterexamples to *modus ponens*. In being convinced of these counterexamples, McGee in effect seemed to reject →ELIM. Surely, Williamson plausibly insists, McGee didn't thereby cease to have the concept of the conditional.

Either McGee is right or he is wrong. If he's right, then he continues to possess the concept of the conditional even by inferentialist lights. We simply tend to *state* an oversimplified Intro rule that fails to apply in his cases. If McGee is wrong, I doubt we face a counterexample to (ii). A defender of (ii) may allow that the way in which one accepts certain rules that are constitutive of certain logical concepts is by having a disposition to deploy those rules in any relevant case. Now take one of McGee's putative counterexamples:

(1) If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.
(2) A Republican will win the race.
_____
(3) If it's not Reagan who wins, it will be Anderson.[144]

McGee accepted (1) and (2) and rejected (3). I take it that most of us would feel the force of this. This *does* look like a bad argument. So we all share the disposition McGee has. So it's not as if McGee's acceptance of rules in the relevant sense *changed* when he saw this putative counterexample. He was *already* disposed to react by accepting (1) and (2) and rejecting (3).

What happened with McGee is that the case led him to be uncertain *about the proper formulation of the rule he was already accepting in deductively presumptive activity with the conditional*. It led him to doubt whether this rule was best captured by →ELIM. Assuming that he was wrong, he had a misleading perspective *on the rule used*. But his first-order dispositions—ones shared by most of us, and in cases like (1–3) where we have McGee's intuition—remain the same. If so, there is no counterexample to (ii). McGee had all the same relevant dispositions as the rest of us, including those of us who try to *describe* our dispositions as dispositions to accept →ELIM. The question is whether we were right in having this perspective on our dispositions and our rule acceptances. The disagreement isn't in our dispositions or the rules we follow. It is in our *description* of them. And it is the rules we *follow* that determine concept possession on (ii), not how we describe them. So Williamson has no counterexample. Even if he *did*, it wouldn't have much of an impact. What is crucial for my argument is that we *do* accept certain Intro and Elim rules, and that when we reason in a deductively invalid way and seem irrational, this irrationality traces to disharmony in the rules we accept. Even if McGee turned out to accept a different rule, the fact that he is disposed to accept *most modus ponens* inferences suggests that there will still be an Intro rule he accepts that is in disharmony with the Elim rule that gets accepted when one applies affirming the consequent in deductively presumptive activity.

3.3.   How epistemic rationality is internalist

I've offered a high-level account of rationality that crosscuts the epistemic and practical domains:

_____

b) and (2008). But his arguments for this are less than convincing. Indeed, they are at odds with the best solutions to some of *his own puzzles* (esp. his radical nativism paradox), as I argue in Sylvan (ms₂).

[144]McGee (1985: 215); I simply borrow the citation from Williamson (2003: n.3).

rationality consists in responsiveness to apparent reasons and requirements of coherence. It is decidedly internalist in at least one sense, as I'll explain now. As I see it, this is no defect. It is what we should expect, given core facts about rationality in the familiar sense—viz., its narrowness, its perspectival character, and the fact that it is subject to act-agent blur.

Of course, this may mandate prying epistemic rationality apart from epistemic justification and truly good epistemic reasons. Indeed, it will do so if one lacks a solution to the problems about the deontic significance of rationality that I will detail in Chapter 3. So, while rationality belongs to (certain) internalists, the victory may be Pyrrhic. To own what *matters*, they must accomplish a more difficult task—one they have, I believe, never even tried to discharge. This is all to be expected. My distinction is standard in the practical literature, and there is a now long-standing and pressing challenge about explaining why practical rationality matters in that literature.

There are many possible forms of internalism about epistemic rationality. I'll focus on the applicability of each to *ex post* epistemic rationality. This is because *ex ante* epistemic rationality raises no new issues when it is viewed—as I've viewed it—as amounting to being in a position to believe with *ex post* rationality. One weak internalist view is:

> **Supervenience Internalism** (SI): If S and S* are nonfactive mental duplicates, then S epistemically rationally has a doxastic attitude *vis-à-vis* P iff S* does too.

This is almost certainly true, I think. Let's see why.

Take the basic case first—the rationality of noninferential beliefs. Here, someone believes rationally only if she directly believes for a sufficient apparent reason in the secondary strong sense. Whether someone has a sufficient apparent reason is determined by the appearances, so that subjects who are "appeared to" in all the same ways will have all the same apparent reasons. Now, if one adopted a *disjunctivist* view about appearance, and claimed that it's veridically visually appearing that X is F *just is* one's seeing X's being F, SI might fail. But while some disjunctivisms are plausible—ones that claim that the *quality* of the reasons differ in good and bad cases—this one is not. Assuming it's false, the *having* of sufficient apparent reasons supervenes on nonfactive mental life. What about the *believing for* sufficient apparent reasons? Well, that is a matter of the appearance that there is a sufficiently good reason playing a key role in explaining why one's motivating reason for belief is what it is. The way this goes is by a certain kind of causal process occurring involving the appearance and the belief—a process that constitutes a *manifestation of one's relevant apparent reasons-sensitive ability*. This will be a mental process, and one non-factive mental duplicates can share: after all, it is plausible that they believe *for the same reasons*, even if the *quality* of the reasons differs. So SI plausibly holds for the noninferential case.

What about the inferential case? Well, in the *non-deductive* inferential case, I cannot see the pressure to invoke anything external to the subject's nonfactive mental life to explain why the facts about rationality. The total ultimate appearance that the rule used is reliable is enough, as I suggested in the last section. One cannot be irrational for using a non-deductive rule that is in fact unreliable, when the ultimate appearances suggest otherwise. Indeed, insofar as one decided to use a rule at all, one would be to some degree irrational in using anything less than the best suggested by the ultimate appearances. So the non-deductive inferential case adds nothing new.

The *deductive* inferential case is the hard one. If the account I gave in the last section is on the right track, what explains why rationality seems more demanding is a modest inferentialism about logical concept possession and identity, together with clearly SI-supporting requirements of rationality we used to explain the facts of rationality in the other cases. These facts, and not facts about the

conditional reliability of the inference *per se*, make rationality more demanding here than it was in the non-deductive case. But these facts supervene on one's non-factive mental life. The fact that one necessarily accepts certain patterns of inference as constitutively valid for the concepts in play and necessarily aspires in a general way to be governed by them in order to even *have* the logical concepts at issue is what generates the basis for the increased demandingness. That is a fact about non-factive mental life—one about the (psychologically narrow) concepts deployed and their possession conditions. A story like this really seems needed to explain the disanalogy between the non-deductive and the deductive case, since *inductive* rules certainly are *not* similarly tied up with the constitution and possession conditions of concepts in a similar way. This story will yield Supervenience Internalism in the deductive case.

Whether epistemic rationality is internalist in *other* senses is considerably less straightforward. Often the term "internalism" is associated—oddly, I think—with various *higher-level requirements*. Some of these are astonishingly implausible, such as the KK principle and the JJ principle. We hardly need Williamson to see that these principles fail! They obviously lead to regress, as sophisticated internalists know.145 Given that we've conceptually separated rationality from other epistemic standings, regress wouldn't clearly be a concern about a *mixed* principle, such as:

> **RPK**: S is rational in believing that P only if S is in a position to appreciate this simply by reflecting on her nonfactive mental states.

RPK should be distinguished from a weaker principle—namely:

> **RFPK**: If F is a factor that bears on whether S would be rational in believing that P, S is in a position to recognize F as such a factor by reflecting on her nonfactive mental states.

I am doubtful about both principles. *Ex post* rationality on our view requires believing for an apparent sufficient reason in the secondary strong sense. That, in turn, was a matter of the appearance that there is a sufficiently good reason playing some role in explaining why one's reason for belief is what it is. While I do agree that the process in virtue of which one believes for an apparent reason in this sense is a mental process, and the ability to implement this process is a mental ability, it does not seem remotely clear that these factors will be *reflectively luminous* in the way required by RPK and RFPK. Nor does the absence of luminosity strike me as rendering rational belief impossible. But it would if these principles were true. This is obvious for *ex post* rationality. And given the fact that *ex ante* rationality on my view requires being in a position to believe with *ex post* rationality, this holds for *ex ante* rationality as well. This is a strike against these principles, not against my view. (It is also a strike against a *stronger* version of Supervenience Internalism the adds "conscious" to "nonfactive".)

Some importantly different principles that are related in spirit that are more plausible. Consider

> **Iterative Appearances Internalism**: It is rational for S to believe that P only if it (ultimately) appears that there are sufficient (ultimately) apparent reasons for believing P.

This is suggested by the ideas that led to our account in the first place together with intuitions about incoherence. Suppose it *doesn't* ultimately appear that there are sufficient ultimate apparent reasons for believing P. After no amount of reflection would it strike you that there are apparent sufficient reasons for believing P. I see incoherence in believing P in circumstances like this.

145Fumerton (1995) rejects the requirements for this reason. He may be the most extreme internalist around.

Interestingly, the subsuming requirement must be narrow-scope. For the conflict state is one of

> (A) it's failing to (ultimately) appear that there are sufficient apparent reasons for believing P

while

> (B) one believes that P.

One can't exit the conflict of (A) and (B) by *forming the appearance* that there are sufficient apparent reasons for believing P. Appearances in the relevant sense aren't things that one can form for (apparent) reasons: they lie outside the scope of rational cognitive agency. This was why it was helpful to use them in retreating from the Pure Doxastic View. While I don't buy everything Kolodny says, he is right to say that if there is a wide-scope rational requirement governing a conflict, it must be possible to exit it by rationally eliminating either one of the conflicting states. Since we can't rationally eliminate (A), eliminating (B) is the only option. So, assuming that there *is* a rational requirement banning the incoherence between (A) and (B), it will be narrow scope. But this is just to say that something even stronger than the contrapositive of Iterative Appearances Internalism holds—namely:

> **Strong Contrapositive-IAI**: If it does not appear that there are sufficient apparent reasons for believing that P, it is rationally required that one not believe that P.

Of course, if it is rationally required that one believe that P, one would be *ex ante* rational in believing that P. So we can simply conclude that Iterative Appearances Internalism is true by weakening the consequent of this intermediate conclusion and contraposing.

But one might wonder at this point whether the coherence side of rationality is pressuring us out of the idea that rationality requires responsiveness to first-order apparent reasons at all. Couldn't there fail to be an ultimate appearance of ultimately apparent reasons *even if* there are some ultimately apparent reasons? But if rationality required one to refrain from believing P when there would be no ultimate appearance that there are ultimate apparent reasons for believing P, wouldn't it simply follow that rationality would require one, in this case, not to respond to the ultimate apparent reasons? What about third-order appearances? Wouldn't the lack of these rationally require one, by the same reasoning, not to believe that it second-order appears that there are sufficient reasons? Indeed, when we start ascending, and allow for gaps between levels of appearance, it looks like we could be rationally required to believe *nothing whatsoever*. Surely that cannot be right! Isn't this a kind of regress argument against Iterative Appearances Internalism, and hence against some of the assumptions that entailed it? Perhaps we were simply wrong to think that there *is* a kind of irrational incoherence between (A) and (B).

One could now reject Iterative Appearances Internalism. I have little at stake in the claim. But there is a tactic. Evan Fales once defended a doctrine on which it's being transparent that P simply entailed it's being transparent that it is transparent that P. He went on to *use* this principle to motivate the JJ principle, given his broader internalist commitments. I think this defense fails, basically for the reasons noted by Bergmann (2006: 38–43). But the originating principle remains plausible, and what Fales said has intuitive force: "Transparency, by its very character must (and does) have a kind of self-sufficiency. To be transparent is to be transparently so; were it otherwise, it would not be a case of genuine transparency."[146] The same remarks might seem plausible for ultimate appearance. If it failed to ultimately appear *that* it ultimately appears that P, how could it ultimately appear that P? To suppose otherwise seems conceptually incoherent. Layers of appearance come cheap as a point of necessity, in much the same way—to borrow Fales's nice analogy—as layers of truth (i.e., $T(P) \rightarrow T(T(P))$). If one has this principle, one can close the gaps

---

[146]Fales (1996: 163–4).

between levels of appearance. With no gaps, bad conclusions don't follow.[147]

In any case, whether we ought to accept Iterative Appearances Internalism is a more complicated matter than it might have seemed. While it may be defensible, it is no solid commitment of mine. Commitment to it certainly doesn't follow from my own account of epistemic rationality.

A final variety of internalism is worth considering. Notice that the version of Supervenience Internalism I embraced earlier included *no temporal restrictions*. It is thus crucially weaker than the kind of claim that most card-carrying internalists really want to defend—namely:

> **Current Time-Slice Internalism** (**CTSI**): If S and S* are nonfactive mental duplicates *at t*, then S would be rational in having some doxastic attitude *vis-a-vis* P *at t* iff S* would be.

Internalism has standardly been a current time-slice position akin to this one, though internalists have claimed to be interested in *justification* and not just rationality in the ordinary sense.

Is CTSI defensible? While I'll express some uncertainty, I think stock objections to current time-slice internalism are less potent when make a distinction between justification and rationality.

To see why, recall an old problem for the current time-slice view about justification—viz., the *problem of forgotten evidence*. At *t*, you acquire clear sufficient evidence for believing (P:) that Samuel Bancroft is responsible for the Delaware Art Museum's having the largest collection of Pr-Raphaelite art outside the UK. Between *t* and a much later time *t**, you forget all this evidence, but continue to believe P. The stock objection to current time-slice internalism about justification then runs as follows. You are justified in believing that P at *t**. You would be considerably less justified in believing that P at *t** if your belief lacked this estimable ancestry, which has, of course, left no traces in *t**. So it follows that some of the factors that contribute to the justification of your belief at *t** lie beyond *t**. Of course, one *might* be tempted to say on behalf of current time-slice internalism about justification that one can continue to appeal to one's current knowledge that most of what one remembers was learned in an epistemically proper way. But as Goldman forcefully noted, this is not enough for extensional adequacy:

> Admittedly, [one] has *some* evidence, but is this evidence sufficient for justification? Surely not. In a variant case, suppose that Sally still has the same background belief—namely, that most of what she learned was learned in an epistemically proper manner—but she in fact acquired her [forgotten evidence] belief from the *National Inquirer* rather than the *New York Times*. […] Then even with the indicated current background belief, Sally cannot be credited with justifiably [holding her

---

[147]This would mandate a different way of responding to the "problem of the speckled hen" than the one I gave earlier. It is one that some people have considered. The idea would simply be to deny that it is apparent, in the sense of "apparent" relevant to expressing the kind of internalism at issue, that the facing side of the hen has whatever precise number of speckles it actually has.

Of course, this is a double-edged sword. The burden is to explain in more detail the sense of "apparent" on which this comes out as true. Some ways of explaining what it would be—ways, say, that invoke, say, a *noticing* of the specific content of one's experience—are unhelpful. Sosa (2003) famously complained that if the *noticing* here is doxastic, then if it is to play any role in rendering belief rational, the constitutive belief itself had better be rational. While the theory would then be plausible, it would be *trivial*.

But I don't buy this complaint. Suppose I am a terrific subitizer. I have inscribed some dots on a certain surface that one of my otherwise reliable and convincing friends falsely tells me will generate an optical illusion of there being N+5 dots. I got a bit lazy as I was inscribing these dots—I was distracted by other thoughts, and thought that I inscribed 10 dots when I actually inscribed 15. Being the great subitizer that I am, it strikes me that there are 15 dots as I look. But I've convinced myself—as I would with a real illusion like the Müller-Lyer case—that this appearance is illusory. Here I notice all 15 *qua* 15, but disbelieve that there are 15. While in normal cases I could add the belief, that would be a distinct matter.

forgotten evidence belief]. Her past acquisition is still relevant, and decisive. At least it is relevant so long as we are considering the "epistemizing" sense of justification, in which justification carries a true belief a good distance towards knowledge.[148]

Can this type of argument be transposed to yield an argument against CTSI?

Not clearly. In our original case, the contents of the total nondoxastic appearances at *t** are *not* going to count as strongly in favor of the content of your belief as they did at the earlier time. Bearing this fact steadily in mind, I don't find it remotely intuitive that you would be even nearly as rational in believing that P at *t**. If you reflected on your belief at this time, you couldn't rationally regard yourself as being in as good a position to continue accepting it as the counterpart of you who had the mass of evidence would have been. The fact about the ancestry of your belief seems to make no difference, insofar as that is beyond one's present ken.

Of course, one can try to appeal to one's knowledge that much of what one remembers was likely to have been acquired in the light of sufficient epistemic reasons. One could claim that this makes one considerably rational after all. There is some promise in this thought, since Goldman's concern does not transpose. If there is a separation of justification and rationality, I do not think that rationality can be understood in some "epistemizing" sense that "carries a true belief a good distance towards knowledge". *That* was the sense invoked in the Goldman quote. Indeed, I'd deny that rationality in the ordinary narrow sense is even required for knowledge, *in part on the basis of forgotten evidence cases*. Rationality can be *knowledge-obstructive*, in addition to being possibly *falsehood-conducive*. So Goldman's argument against appealing to this kind of background belief to ground sufficient justification won't transpose for sufficient rationality.

But background knowledge alone will not be enough in *all* the cases where we *can* reasonably claim there to be justified belief. Sosa (ms) points to a case that illustrates this:

> At noon on a certain date you are mistreated as a child. […] You store that belief for months, even years. Maybe you retain it through excellent memory. In general people would not remember so well. Maybe in general your own memory does not work so well. But it does in this case, on this sort of subject matter. That event stands out in your mind, and your memory of it is outstanding. The perception-plus-memory manifest in your continuing belief is of the highest quality. When perception and memory work as they do in this case, they are *extremely* reliable. Compatibly with that, your second-order competence can decay. Just based on common sense, you may come to doubt that your memory of that event is as good as it in fact is. […] Human beings do not recollect as reliably as had been thought, especially not on *such* subject matter. By hypothesis, however, *your* memory *is* in this case extremely reliable.[149]

Here the background belief doesn't help. While it is not implausible that one can be justified in the "epistemizing" sense in this case, I do not agree that one would be rational in continuing to believe. Plausibly, this flows from the account of rationality I've given, since the ultimate appearances do not provide sufficient apparent reasons for the belief. This strikes me as a not implausible verdict if we are self-consciously understanding rationality in the familiar sense. So the defender of CTSI can, I think, simply bite the bullet, and note that it is a soft one. One will be irrational in this case and others where the background knowledge cannot make a sufficiently clear contribution to the apparent reasons now in favor of your belief. This implies nothing about whether one is justified or would know, given the fundamental disconnections in play.

It is worth adding that the disconnect between rationality and justification saves CTSI from a

---

[148]Goldman (1999: 280–1).
[149]Sosa (ms: 7).

different class of forgotten evidence cases that are even more acutely pressing for current time-slice internalism about justification. These are forgotten *negative* evidence cases. Consider a variation on a case from Goldman (2009: 16). At t, Ursula acquires by testimony a justified belief in a certain generalization G. Between t and and a later time t^*, Ursula acquires a mountain of counterevidence against G, ignores it, and simply retains her belief. By t^{**}, which is significantly later than t^*, she has forgotten all the counterevidence, and has, indeed, acquired new *positive* evidence. So, in the light of the total evidence accessible to her at t^{**}—which, let's stipulate, simply coincides with the total ultimate nondoxastic appearance—she would *appear* to have sufficient reasons to believe G, assuming she takes any doxastic attitude toward G at all.

I share Goldman's intuition that Ursula wouldn't be justified in retaining her belief in G at t^{**}. But there is a simple argument that she is rationally required to believe G, assuming that she is interested in the question. Assuming that she is interested in the question whether P, disbelieving P and suspending judgment on P would be irrational, given the verdict of the total nondoxastic appearances at t^{**}. So believing P is the only alternative, assuming she takes a stance. Indeed, this is exactly the argument that Feldman gave in reply to Greco when the latter advanced a structurally analogous objection to Goldman's new forgotten evidence case.[150] Of course, as Jackson (2011) correctly insists, it is a huge step from Ursula's being rationally required to believe G, given her interest in the question, to her being justified in believing G. But since it is rationality that interests us now, we ought to mimic Feldman and embrace the conclusion.

These cases also illustrate something about the relationship between rationality and unqualified hypological epistemic properties like *being blameless in believing*. At the end of Chapter 1, I noted that hypological properties do not generally supervene on the current time-slice, given the distinction between *primary* and *secondary* hypological evaluations. Forgotten negative evidence cases are a fine illustration of this fact. As Greco (2005) notes, it is pretty intuitive that that the subject is derivatively irresponsible and hence *not* blameless in holding the belief at the later time in these cases. That is my intuition about Ursula. This shows that we can be rationally required to believe things that we would be to some degree epistemically blameworthy in believing.

This is, as it happens, a generally plausible point about rationality. Consider a practical example. At t, you are told by a reliable authority that giving a certain potion to your best friend Jones will cure all his ills. You form this belief, and form the intention to obtain the potion to give to Jones. Between t and t^*, you are in a position to acknowledge some further *conclusive* evidence from an *even more reliable* authority that indicates that the potion would kill Jones. But you are so caught up in other projects at the time that you simply fail to heed this evidence. At t^{**}, you've lost access to the counterevidence, and acquire some new misleading evidence that the potion is even more effective in helping Jones than you initially thought.
If you had remembered the old evidence, your total evidence would not have been sufficient, and you would have been rationally obliged to revise your belief and intention. But given where you are now, you'd be irrational in doing anything other than giving Jones the potion. By current total appearance, failing to get the potion to him would result in his death! Still, on many plausible theories about the relationship between culpable ignorance and later intention, you would be derivatively blameworthy in giving Jones the potion. So you can be rationally required to do what you would be blameworthy in doing. Separating rationality and justification isn't *all* we can reasonably do. We can also reasonably separate rationality and pure hypological evaluations of praiseworthiness and blameworthiness, either of an epistemic or a practical sort.

In any case, the core upshot is this: one of the main concerns that people have had about current time-slice versions of internalism do not extend to undermine CTSI. Indeed, thinking correctly

---

[150]See Greco (2005: 260–262) and Feldman (2005: 282) for this exchange.

about these cases *supports* CTSI. Ought we, then, to accept CTSI?

I'm uncertain. For one thing, we have to be careful about distinguishing *ex ante* and *ex post* epistemic rationality in this connection. *Ex post* epistemic rationality involves believing for sufficient apparent reasons. As I conceded at the outset, it takes a *process* to enable one to get connected with sufficient apparent reasons in the relevant sense. Insofar as one *now* believes for sufficient apparent reasons, that is a fact that partly owes to facts outside the current time-slice. So there's a clear sense in which the facts that make claims about current *ex post* rationality true lie beyond the current time-slice. But this does not trivially suffice to refute CTSI. After all, believing for a sufficient apparent reason is not *itself* a process: it is the result of a process. All we've shown is that it takes a process to *get* us to believing for a sufficient apparent reason. Why can't one take me *now*, as I *now* believe for a sufficient apparent reason, (nonfactively) mentally duplicate me, and end up with another subject who believes for a sufficient apparent reason? Unless the dependence of believing for a sufficient apparent reason on a previous process is the kind of *constitutive* dependence that holds between a mental state with wide content and the environmental factors that ground that content, we do not yet have a counterexample to CTSI.

There is another concern one might have. Both *ex post* and *ex ante* epistemic rationality turn on how things are *relative to the total ultimate appearance*. But the total ultimate appearance blends in the upshots of stored and dispositional appearances. As I said before, this ultimate appearance at *t* has it that P only if it's true that *if* all stored and dispositional appearances existing at *t were* made occurrent, P would be true or likely enough conditional on the total content of the resulting ultimate appearance. Don't we then face Goldman's problem of concurrent retrieval?[151] Transposed in our vocabulary, the problem would have to be this. Sometimes only certain conjunctions of stored mental states can provide sufficient apparent reasons. Some of these conjunctions may be too big to be concurrently retrieved. So it may be psychologically impossible for one to have them all in view at once. Yet surely these states still collectively play a role in making it *ex ante* or *ex post* rational for one to believe, disbelieve or suspend judgment.

That is true. But I don't see that the reverse follows from our view and CTSI. Nothing in our view plus CTSI implies that one has to have all these states in view at once to believe for the ultimately apparent reasons they collectively give, or for these ultimately apparent reasons to make certain beliefs *ex ante* rational. All these two views imply is that facts about epistemic rationality at *t* supervene on truths about the ultimate nondoxastic appearance at *t*. Yes, that is a partial construction out of stored and dispositional mental states. But it is a counterfactual construction: P is ultimately apparent at *t* only if it's true that if certain stored and dispositional appearances *existing at t were made occurrent*, the total resultant appearance would be such that P is likely conditional on its content. Why not say that one can at *t* believe for the sufficient ultimately apparent reasons partly given by *prima facie* appearances while these are appearances are stored or unmanifested? Neither my view nor CTSI says otherwise. If not, it is also perfectly possible on this conjunction of views for the stored or dispositional *prima facie* appearances to play a role in making it *ex ante* rational to believe, disbelieve or suspend judgment.

Remember that Goldman's problem was a problem for views that put emphasis on current *conscious* mental states. The analogue of the view he was attacking is a view on which some factor can only play a role in determining whether it is rational for someone to believe something *if that factor is conscious or can be gotten into the "specious present" of consciousness surrounding the current time t*. What we really see is that the problem is not a problem about current time-slice views *per se*, but rather about insisting that rationality-making factors must be *all* consciously accessible within the current time-slice. But the things that make up the total ultimate appearance at *t* were never assumed to all be

---

[151]See Goldman (1999: 281–2) for the version that targets current time-slice internalism about justification.

conscious or even retrievable concurrently in point of real psychological fact. The reverse can be assumed, since the total ultimate appearance was simply to be a counterfactual construction from the nondoxastic appearances existing at t, stored, dispositional, and occurrent. The possible worlds that make the counterfactuals true may be distant.

That doesn't prevent this concept from playing a key theoretical role. We just do have a notion of how things look from the point of view of all S's presentational mental states, stored, occurrent and dispositional. It captures the relevant perspective for limning the perspectival character of rationality. It needn't be S's own *psychologically possibly <u>conscious</u> perspective*. Even when conjoined with CTSI, my view needn't imply otherwise. CTSI implies that rationality at t turns on nonfactive mental facts at t, not *conscious* or *collectively consciously accessible* ones.

So CTSI may be a truth about epistemic rationality. Given what I said about forgotten evidence cases, what makes it an *interesting* thesis and makes room for bold disconnections between justification and rationality survives. I'll exploit some of these features of epistemic rationality in the next chapter to formulate some problems about the deontic significance of epistemic rationality. But because I have a lingering worry that some descendants of Goldman's objections to time-slice views about justification may arise even for a time-slice view of rationality, I'm wary about building endorsement of CTSI into my picture. It certainly doesn't *follow* from my picture.

III

PROBLEMS OF DEONTIC SIGNIFICANCE

> The requirements of rationality may be, in part, standards of appraisal, by which we measure how far someone manifests a kind of virtue or proper functioning. But they also seem to call for certain responses. […] [H]ow can this be, if we don't have reason to comply with rational requirements?
>
> —Niko Kolodny[152]

---

[152]Kolodny (2005: 513).

Sadly, this is all pie in the sky. […] It does not follow by means of any inference I can find that rationality is normative.

—John Broome[153]

1.   *Introductory remarks*

The aim of this chapter is to generalize to epistemology a problem about the deontic significance of rationality that has become pivotal in recent literature in the philosophy of practical reason. Once epistemologists accept the distinction between rationality and responsiveness to (possessed) genuine reasons, some central internalism/externalism disputes are superseded by this problem, with the burden falling on the internalists. While many internalists are probably right about epistemic rationality, the victory is Pyrrhic: a deeper worry remains about epistemic rationality's deontic significance that is structurally like a well motivated pessimism about practical rationality. Being an internalist about *justification* at minimum requires a response to this. Given that the increasingly orthodox view in the practical literature *is* pessimistic, the burden is hefty. Indeed, there will, we'll see, be a broader burden on anyone who thinks perspective on what epistemic reasons we have can interestingly determine as such *how good* the epistemic reasons for us are *ultima facie*. So there are implications for externalist views too—e.g., most externalist views that incorporate "no defeat" clauses while understanding defeat *very liberally* (and so many views *simpliciter*), and all views that allow some kinds of putative higher-order evidence to determine what to believe at the first order. So the problems matter a lot.

The originating problem in the practical literature is often called the "normativity problem". The literature on it often comes under the heading of the "normativity of rationality". I do not want to follow this nominal tradition. I prefer to use 'normativity' broadly, and to see normative domains as being split up as suggested in Chapter 1 into three subdomains: deontic, hypological, and evaluative. I prefer doing things this way mostly because using any other term so broadly does violence to it. And there are already many who use 'normative' as the broadest term. It is crucial to realize that *nobody* in the recent literature questions whether rationality might crucially determine certain *hypological* assessments or certain narrow *evaluative* assessments. Niko Kolodny, for instance, explicitly allows that rationality might be "normative" in these other ways. One can see this in the epigraph. He just found it is puzzling how this could be the whole story, and so was compelled to give an *error theory* to soften the intuitive blow of his conclusions. The problems that have arisen are about the *deontic* significance of rationality—about whether we can have *strong genuine reasons* in any case simply for being rational *per se*, or about whether we *really ought* to be rational *as such*. While this shows the core problem to be narrower than it might sometimes sound to outsiders within earshot of these discussions, it remains as significant as ever. For many—particularly internalists in both epistemology and the philosophy of practical reason—have either claimed or presupposed that rationality as such is deontically significant. (There were even days within memory when rationality was *the paradigm* of deontic significance.)

As in the last chapter, I am going to start with a substantial discussion of the practical sphere. The problems have mostly been discussed in this connection, so familiarizing the reader with them requires discussing these standard problems first. But another reason is that I have a special obligation to show that the problems about deontic significance *still exist*, given that I've defended a *much less narrow view* about rationality than the one accepted by many who care about these problems. It will be crucial for the aptness of my analogies that the problems about deontic significance *survive* the transition to my less narrow view. I think they do. But this is not a trivial

---

[153]Broome (2008: 175).

claim. There *are* people in ethics who think that the *ought* that matters for deliberation—and that is connected with *possessed reasons*—is the evidence-relative 'ought'. I'll need to consider whether the best reasons for thinking that the evidence-relative 'ought' *is* deontically significant are also reasons for thinking that the 'ought' of rationality—understood as turning on the balance of *apparent reasons* in the nondoxastic sense—is deontically significant. As I'll argue, they aren't. So the core problems about deontic significance remain when we adopt my less narrow view. Independent solutions are required *beyond the right change of view about the nature of rationality*.

1.1.   *Bootstrapping and implausible conflicts in the early literature*

I start with a quick history of how the problems of deontic significance originally came to light. In doing this, I will spell out what I think the heart of these problems really is, so that I can make clear how and why they generalize even for views that depart from some of the structuring assumptions of the originating literature (e.g., the Pure Doxastic View refuted in Chapter 2).

In his (1999), Broome held that norms like the following should be read as wide-scope oughts.

> **Enkrasia**:   Rationality requires that if S believes she ought to A, S intends to A.

The wide-scope 'ought' formulation of Enkrasia is:

> **O-Enkrasia-WS**:   S ought to satisfy this conditional: if S believes she ought to A, she intends to A.

Why accept this reading? Well, consider the narrow-scope 'ought' reading:

> **O-Enkrasia-NS**:   If S believes she ought to A, she ought to intend to A.

If this reading is true, we end up with a crazy kind of bootstrapping. Simply believing that you ought to A will make it the case that you ought to A. This is obviously wrong. Normative beliefs are not self-verifying. So, since Broome initially simply assumed that we really ought to be rational, he concluded that the correct reading of Enkrasia must be O-Enkrasia-WS. For there are *two* ways to comply with that 'ought': dropping the normative belief, and adopting the intention. On the face of it, then, no bootstrapping arises for the wide-scope 'ought' reading.[154]

Broome also thought that 'rationality requires' couldn't be *seriously reason-implying* if it is read as having narrow scope. Crucially, I think this is equally compelling. The idea that a serious normative reason that supports intending to A just pops into existence simply in virtue of a false normative belief about A-ing which then can substantially compete with the other reasons bearing on whether to intend to A is implausible. How could an *illusion* of deontic significance generate some real and serious deontic significance that competes with antecedently existing deontic significance, and which might in principle defeat that antecedently existing deontic significance? That idea seems crazy, even if it is not as vividly crazy as the problem about self-verification that arises if we take 'rationality requires' to be *ultima facie* 'ought'-implying. Perhaps some *very weak* reason pops into existence. But too weak, surely, to have any effect. Too weak, more crucially, to help explain why we ought even *prima facie* to be rational as such.

So the problem isn't fundamentally about self-verification. It is about the implausibility of thinking that our normative beliefs *vis-à-vis* A-ing or intending to A can simply pop into existence new seriously weighty reasons that will compete with the reasons that were already out there bearing on A-ing or intending to A. More vividly and accurately, it is about the implausibility of thinking

---

[154]Some have argued that even Broomean wide-scopers face a problem of bootstrapping—e.g., Raz (2005), Setiya (2007) and Schroeder (2009). But as I argue in Sylvan (ms4), this isn't really (interestingly) true.

that what would antecedently be an *illusion* of deontic significance could *create* substantial *real* deontic significance—that a *false impression* that there are decisive (or sufficient) reasons (not) to A could create *truly strong reasons* for (or against) A-ing that would compete with the strong reasons that are antecedently lined up against (or for) A-ing. This is what I take the **core problem of bootstrapping** to be. One core problem of deontic significance, as we'll see, lies in the thought that *if* rationality is seriously deontic, then bootstrapping ought to be possible.

Obviously, Broome's argument for wide-scoping 'rationality requires' only works if we *presuppose*

> that the 'ought' implicit in 'rationality requires' is the same ought that shows up in the content of S's belief,

and

> that the 'reason' that we assume (*arguendo*) to be implied by the requirement is a reason of the same kind as the reasons for A-ing that S is already thinking about.

In the first wave of the literature, people agreed that insofar as rationality has deontic significance, it is going to be *because* the oughts and the reasons that it gives rise to are the *same* as the oughts and the reasons that *really matter*, and that *we aim to comply with*. I think this is a good assumption to make. Simply thinking that there are several incommensurable oughts or kinds of reasons does nothing to resolve the mystery. It *enhances* the mystery, as Kolodny noted:

> This […] also because of the difficulty of understanding how we are to be governed by these two autonomous 'ought's. As we have seen, what one is rationally required to believe or intend will sometimes conflict with what one in fact has conclusive reason to believe or intend. If reasons and rational requirements give rise to autonomous, primitive 'ought's, then what *really* ought one to do in such conflicts? Some compromise between what one has reason to do and what one is rationally required to do? Or is it rather that we cannot even ask what one *really* ought to do, only what one ought-in-the-reasons-sense to do and what one ought-in-the-rationality-sense to do? These are unpalatable alternatives.[155]

What other theorists in the first wave of literature after Broome then became skeptical about was the early Broome's further presupposition that requirements of rationality really do generate oughts or reasons in the fundamental sense that really matters.

Why? Well, some of the early skepticism arose in large part from the recognition that wide-scoping across the board is untenable. Kolodny argued against Broome's far-reaching wide-scoping program in "Why Be Rational?". He argued that Enkrasia *is* a narrow-scope principle, and on this basis rejected Broome's further presupposition that the 'ought' of rationality is the fundamental 'ought'—the one that really matters. It is at best just the deontically *pallid* kind of 'ought' that shows up with *any* system of requirements or standards of correctness—including, say, the requirements or standards of *etiquette*. Kolodny agreed that Broome's presupposition of deontic significance together with a narrow-scope reading of Enkrasia would lead to an implausible kind of bootstrapping. He simply took this to be a *reductio* of Broome's presupposition! While Broome continued to defend wide-scoping in later work (e.g., Broome (2007b) and (ms)), he ended up *joining* Kolodny in being pessimistic about the deontic robustness of requirements of rationality. (See for instance Broome (2008), from which I get my epigraph.)

1.2.   *Bootstrapping and implausible conflicts transposed for appearance-relative requirements*

---

[155]Kolodny (2005: 556).

Now, in the last chapter, I expressed doubt about Kolodny's main arguments for thinking that requirements like Enkrasia are narrow-scope. Still, in saying this, I assumed that there are *some* narrow-scope requirements of rationality. I just disagreed about *what* they are. Indeed, I think there *must* be some such requirements to avoid too much symmetry. Simply adding another wide-scope principle would not be enough to break the symmetry; it would, indeed, just create more symmetry.[156] So I think we cannot live without some narrow-scope requirements of rationality. My replacement of the narrow-scope reading of Enkrasia was:

> **Appearance-Relative Enkrasia-NS**: If the total ultimate appearance for S has it that there are decisive reasons for her to A, S is rationally required to intend to A.

What we have to ask is whether the reasons that led people to be skeptical about the deontic robustness of the *belief-relative* narrow-scope requirements—e.g., ones we'd get by replacing talk of total ultimate appearance with talk of belief in my requirements—generalize for mine.

I think they do. Remember that the core of the problem of bootstrapping was just the simple thought that it is preposterous to think that an *illusion* of genuine deontic significance could create *serious further genuine deontic significance* that bears on intending to A beyond the kind that existed before the illusion was there, and that could in principle compete with and defeat the reasons that existed before the illusion cropped up. But appearances—even ultimate nondoxastic appearances—can intuitively be just as illusory with respect to the normative facts as beliefs. If that is right, the fundamental concern about bootstrapping can just be transposed if we switch from a belief-relative to a nondoxastic appearance-relative view of rationality.

Here is a more official way to put the point. The following is obviously true:

> *Content Switching Thesis*: For any S who is in a problematic bootstrapping case owing to her normative beliefs B_i with contents C_i, there is another possible S* who differs from S in the following minimal way: instead of having beliefs B_i with contents C_i, S* has total ultimate nondoxastic appearances A_i with the same contents C_i,

The following is also very intuitively compelling:

> *Transposing Thesis*: An ultimate nondoxastic appearance can be misleading with respect to the real normative facts—i.e., about whether there are (strong / sufficient / decisive) reasons to A—in the essentially same way that a mere belief can be.

It is hard to see how *if* a mere belief is misleading with respect to the normative facts, an ultimate

---

[156]Wide-scopers try to address the worry about symmetry by adding further requirements of a rather different type. Clearly, *one* of the things that rationality ought to ban is reasoning from the fact that one has a certain intention to the revision of one's normative beliefs. Broome (ms) and Way (2010) both suggested adding basing principles—or "anti-basing requirements", in Broome's wording—such as:

rationality requires not forming a belief that one has sufficient reasons to A *for the reason that one intends to A*.

While these are probably requirements of rationality, they do not capture all the desired asymmetries. If one believes on clearly good evidence that one's belief that one has decisive reasons to A is false, it would not be rationally permissible to exit the conflict state by forming the intention. Yet one could *arbitrarily drop* the higher-order belief, or *ignore* the clear evidence, or *reconsider* it and regard it as unclear without violating any basing principles *or* some wide-scope requirement governing the relation between the higher-order belief and the first-order belief. Indeed, if Holton (2009) is to be believed, the last of these is closer to what happens in cases of rationalization that we regard as irrational. It is hard to see how to accommodate these points without accepting some narrow-scope principles.

nondoxastic appearance with the same content could necessarily be (significantly) less misleading, just by being an ultimate nondoxastic appearance rather than a belief. The only way this could be true is if the appearance of a normative truth *vis-à-vis* A-ing could itself—just in virtue of being such a thing—constitute a new serious reason bearing on whether to A.

It is hard to see what could motivate such a claim. People can differ vastly with respect to what ultimately nondoxastically appears to them to be the case, particularly with respect to the reasons. This may owe to *substantive defects* on one side. A sociopath may never be struck with remotely correct moral intuitions in the way in which we are, even if he comes to understand popular normative thinking, and forces himself to conform for egoistic reasons. Sure, because nondoxastic appearances do not reflect the person *qua* agent in a way that his beliefs do—they strike him "from the outside", and never as a response to reasons—we cannot *fault the person* for these defects in the way we could fault him for his beliefs or choices made on their basis. But why can't they be defective all the same—just as defective, from a God's-eye point of view, as the beliefs?[157] As I said in Chapter 2, the defects wouldn't be defects of rationality. Rationality turns precisely on responsiveness to apparent reasons in this sense. Nondoxastic appearances aren't the sorts of things that can be responsive to reasons, apparent or real. But that is the point. It is the gap between rationality and responsiveness to real reasons that generates the problems of deontic significance. Switching to the Weak Appearance-Relative View does not bridge this gap.

So there is a burden of proof on one who thinks that the deeper problems behind the kind of bootstrapping that led Kolodny to pessimism—especially the worry about how false normative beliefs can be seriously reason-implying—are solved by switching to a plausibly nondoxastic account of rationality. She will have to explain why misleading appearances *vis-à-vis* what reasons exist by their nature can't be just as out of touch with normative reality as beliefs can be.

Now, it would be unfair to simply presuppose that there is no way of dispatching this burden. The best way would be to consider what is going on in cases where our lack of contact with the reason-giving facts does seem to have an effect on the 'ought' that matters. Many people in the broader ethics literature have thought that the 'ought' that matters can be affected by ignorance. Many think that it is evidence-relative. While I don't think the best account of rationality is rightly viewed as an *evidence-relative* one without a lot of further work—basically, without a solution to the epistemic analogue of the problem of deontic significance—one might be optimistic that this literature will show how misleading normative appearances can also influence the 'ought' that matters for deliberation, as well as what reasons we *have* or their balance.

Let's consider this thought. As we'll see, it is ultimately unpromising.

---

[157]Couldn't the appearances create *some* reason, and so be *less* defective? Perhaps. Perhaps a little less defective. But even if that is so, it does little to show that rationality is *interestingly* deontically significant.

    Notice that I am not hereby suggesting that *truly good evidence that one has reasons* couldn't constitute a serious practical reason. For I argued in the last chapter that mere nondoxastic appearances of reasons are not plausibly *ipso facto* truly good evidence of the existence of these reasons—at least nowhere near sufficient evidence, even sufficient *possessed* evidence. How could it be otherwise, given that we can easily imagine cases like the one I'm imagining, and that there is no plausibility in the thought that the sociopath is hereby simply licensed to act against the balance of possessed or simply existing moral reasons?

    I suppose one could claim that he doesn't really have these reasons. This is not very plausible. Even if it were plausible for some sense of "have reason", there would simply then be a problem of deontic significance for whatever sense of 'ought' is connected with "had reasons" in this sense. The mere existence of some new sense of 'ought' is, of course, obviously not enough. We could use 'ought' to mean 'ought conventionally' or 'ought by the lights of etiquette' in some contexts. In any case, it is just plainly false that, if not for the egoistic reasons, the sociopath would be permitted in the sense connected with the real 'ought' to kill random people when he feels like it.

1.3.   *One way for ignorance to be deontically significant*

It is worth noting at the outset that many cases that naïvely motivate thinking that ignorance can affect what one ought to do actually do nothing to decide between this view and a view on which ignorance can affect only what one would be blameworthy or praiseworthy for doing. Consider, for instance, the following case, which I take from Zimmerman (2008) with slight modifications:

> *Case I.* It would be rational for Jill to believe that giving John drug B would cure him partially, and that giving him no drug would render him permanently incurable. It would also be rational for Jill to believe that giving John drug C would cure him completely and that giving him drug A would kill him. So Jill decides to give John drug C. Alas, it turns out that the appearances were misleading. Giving John drug C ends up killing him, and moreover was objectively the most likely to kill him of the three drugs.

In this kind of case, it is perfectly open to say that while Jill is *blameless* in doing what she does, she does the wrong thing. Further reflection makes this extremely plausible. Giving John drug C would almost certainly kill him as a matter of objective chance. How could doing what is most objectively likely to kill someone be the *right thing to do*? Of course, there may be a competing feeling—viz., that "there is a sense in which Jill ought to give John drug C". But this is feeling accommodated by saying that Jill *ought rationally* to give John drug C and is *blameless* in doing so *for that reason*, even though she really ought not. Whether this has any implications for what genuine reasons there are for Jill to administer drug C is what is in question. Pointing out that there is an 'ought' of rationality does nothing. It makes no difference if you want to call it a "subjective (moral) ought", and say there is a theoretical role to be carved out for this. What we are interested in is the genuine deontic significance of such oughts, and whether they *coincide* with the oughts that *really matter*. Conceding that they do not coincide is to give up the game.

These points support a principle I call **Austin's Razor** after the author of "A Plea for Excuses":

> If   (i) there is an intuition that an agent's act is right [wrong] that can be explained by supposing she is merely blameless [blameworthy] in doing the wrong [right] thing,
>
>   (ii) there is also a clear competing intuition that she *really does the wrong* [*right*] *thing*,
> and
>   (iii) there is a clear basis for excuse [or "accuse", to use Zimmerman's neat word[158]]
>
> then   go ahead and explain away the intuition by drawing a hypological/deontic distinction.

Austin's Razor allows us to simplify our normative thinking and reconcile our intuitions without inviting further puzzlement. Instead of multiplying oughts that must be deemed mysteriously incommensurable or giving up on such attractive thoughts as that it *wasn't really right* for Jill to do what had the *greatest objective chance of leading to John's death*, we can distinguish between a hypological assessment of a person and a deontic assessment of an act. This is very attractive.

Alas, not all cases that feel superficially like *Case I* are so easily sliced up with Austin's Razor. For sometimes clause (iii) is not satisfied. Consider the following case, which I again take with irrelevant alterations from Zimmerman (2008: 17–8), who was inspired by Jackson (1991):

---

[158]Zimmerman (1997) coined 'accuse' as the negative analogue of 'excuse'. Just as there can be *excusable wrongdoing,* Zimmerman claimed that there can be "*accusable*" cases of *permissible action*—cases where one does nothing really wrong, but is nevertheless blameworthy. This is often the most plausible way to think of acting with bad motives in cases where so acting would have no bad effects or really *risk* any bad effects.

Case II. Jill knows that giving John drug B would cure him partially, and that giving him no drug would render him permanently incurable. But as things stand, it would not be at all rational for Jill to settle on a verdict about whether it is drug A or drug C that would cure him completely or kill him. She does know that one would cure him completely, and that one would kill him—just not which one. Knowing that she knows this, Jill picks B.

In *Case II*, it is again *prima facie* plausible that opting for any other drug would have been wrong. So far, so indecisive, given Austin's Razor. But there is a crucial difference. Jill is in a position to know that this is *not the best thing to do*. For she can know that doing either A or C would be better. This is unlike in *Case I*, where by stipulation it was most rational for her to think that she was doing the best thing, and she couldn't see that appearances misled. In *Case II*, she knows that one of the other drugs would cure him completely. She is failing to cure him completely. Typically, when one wants to explain away the intuition by appeal to excusability, the excuse is precisely that the agent was rational in thinking that she was doing the best thing or that she was blamelessly unaware that there was a better alternative. No such excuse is available here.

So the original intuition cannot be sliced into two with Austin's Razor. If she does act rightly, then she acts in accordance with the reasons. Yet how could she have most reason to do what is clearly not best? There is, I believe, a story to be told about this. It is one that works even without making the ought that matters for deliberation *interestingly* subjective or evidence-relative in the way some have assumed. Ignorance can, I'll agree, sometimes have a real effect on what one has most truly normative reason to do, by having a real effect on the relevant objective facts.

To bring this out vividly, consider a further case from Jacob Ross:

> *Three Envelopes.* You must choose one of three envelopes. You are told (correctly, and by a reliable person) that there is $900 in the first envelope. You are also told that there is $1000 in either envelope two or envelope three, and that the envelope that doesn't have this money is empty. You listen to what you are told, and decide to choose the first envelope.

There is a simple explanation not just of why you can rationally pick the first envelope, but also of why you have most reason to pick it. Consider the following facts that hold in *Three Envelopes*:

(i)   If you did not pick the first envelope, you will pick objectively randomly and so be as objectively likely to come up empty-handed as to come up with the $1000;

(ii)   The difference $900 and $1000 is not a great difference.

Facts (i) and (ii) are real facts that constitute a strong reason to stick with the first envelope in your circumstances. Crucially, (i) owes to your witting informational limitations. You are in a position to recognize that (i) and (ii) hold. Of course, you would not have the reason given by (i) and (ii) *if you knew all the facts*. This is unsurprising, since (i) would not longer be a fact. That does not imply that (i) and (ii) do not give you a *truly normative reason* but only something *second-rate*, *fake* or *merely subjective*. You get a truly normative reason that you just happen to have only when you don't know all the facts: one that owes to the *objective chances of success* if you opt for something other than the first envelope, given your complete cognitive state.[159] It outweighs the reason against choosing

---

[159]Dancy (2000: 69) made something close to my point: "[A]ssessing rationality relative to incomplete information does not require us to think that the rationality we are assessing is subjective in any damaging sense […]. It is worth remembering in this connection that the fact that I do not know something can itself be a reason. Suppose that my information is limited, that I know this, and that I have no time to make further enquiries. The action I have most reason to do might not be the one that I would have had most reason to do if I had had time to find out more of what I needed to know. This should remind us of Prichard's example of slowing

the first envelope flowing from the fact that it is not the best option.

This should not be mystifying. The character of one's epistemic position can have genuine and not merely apparent effects on whether one would be successful in intentionally acting. Such effects are real facts about the world like any others. They are not illusions. And they give one truly normative reasons. Being aware of these or being in a position to know them makes it *right* to act in certain ways in which it would not be right to act if you knew everything.

Here it is easy to get confused, since the facts that are providing these truly normative reasons owe to your imperfect mental state. (i) is true because you don't know what the better option is, and aren't in a position to know it. But this does not make it constitute less than a truly normative reason. It is not as if it is some mere phantasmagoria.

It is worth thinking of the effects of epistemic positions in these cases on what one has truly normative reason to do by analogy with the effects of limitations on one's *capacities as an intentional agent*. Indeed, this needn't be a mere analogy. Our case is plausibly subsumed by that case. We *already knew* that you sometimes do not have most truly normative reason to bring about the best state of affairs—namely, in the cases in which you are unable to intentionally bring about the best state of affairs. The reason is given by a fact owing to you: if you tried, you'd objectively probably fail, and not bring about that state. Clear-eyed ignorance can lessen your intentional capacities as much as non-mental handicaps. On reflection, this is obvious. If I blind you and make you deaf, you will probably be a less effective intentional agent, barring clairvoyance, blindsight, or stubborn luck. We ought to think of clear ignorance in many cases in similar terms: a kind of blindness or deafness to the facts. You may know the generic fact that there is an option that is better than all the others. But if you are blind to which option it is, and wittingly so, it need not follow that you have most truly normative reason to try to pursue it. For you might just as likely—objectively speaking, relative to the real chances—fail in trying.

Something analogous, though less extreme, explains the intuitions about *Three Envelopes*. This does not show that truly normative reasons are determined by subjective factors in a *surprising* way. Not any more than familiar observations about the effects of limitations on one's intentional capacities do—observations that obviously arise in thinking about cases where, say, unlocking a safe would bring about the best outcome but you don't know the combination.[160]

1.4. *The best explanation of the real effects doesn't generally help with problems of deontic significance*

It is crucial to appreciate the limitations of this explanation, and what it doesn't show about the

---

down at a junction. In that case, the objectivist seemed perfectly able to say that the main reason for slowing down is that one does not know whether any traffic is coming. If one had known that there was no traffic coming, perhaps one would have had no reason to slow down."

[160]Howard-Snyder (1997, 2005) made something like this point. She went too far, though. She tried to argue against "objective consequentialism"—roughly, the view that we ought to do the best of those acts available to us—on this basis. As Moore (2007) points out, this can be reinterpreted as a point about the relevant sense of "available". See also Zimmerman (2009: 132–8) for a fine discussion of the relevant sense of "available" that is made in full awareness of Howard-Snyder's points.

   Oddly, Zimmerman fails to realize that this very point can be used against his own evidence-relative account of moral obligation, for reasons that will become very clear soon in the main text. This is probably because the objectivist view he explicitly opposes is the view that we ought to do the *best act available to us* (without violating deontological constraints), rather than the view that we ought to do what *the genuine reasons favor* (rather than the reasons that are (epistemically) likely to exist, conditional on the evidence). Still, it's a huge jump from the failure of *this* objectivist view to the truth of his evidence-relative, "prospective" view. This is for precisely the reason that there is a different view from his that accommodates Howard-Snyder's point by seeing it as one about *what genuine reasons are possessed.*

dependence of normative reasons on facts about epistemic positions. It was a key part of the account of why you ought to do what you'd do in *Three Envelopes* that

> (i)   if you did not pick the first envelope, you will pick objectively randomly and so be as objectively likely to come up empty-handed as to come up with the $1000.

No analogous claims hold in all or even many cases where the reasons that *ultimately appear* to exist sufficiently support or prohibit an act. So they cannot be used to motivate an appearance-relative view of *wrong/right* or *reason-supported/-prohibited action* as opposed to an appearance-relative view of *rational/irrational action.*

To make this vivid, consider the following case. You were planning to play Russian roulette on Bill. Unbeknownst to you, he anticipated your act and removed all the bullets. You trap him. You tell him that he is "in for the game of his life". You put the gun to his head. You rationally believe that it's objectively likely that you'll kill him. Far from it! He knows this, and is laughing inside, thinking you a sucker. While he can complain about your earlier *conspiracy*, it is astonishingly implausible to think that he can complain much about what you are doing now, except insofar as it is indicative of your *bad character* or *motives,* or has what Scanlon (2008) called "predictive significance" *vis-à-vis* your later behavior. How could there exist conclusive reasons against pulling the trigger of an *empty gun*, and thereby *risking nothing*? We ought to think that there couldn't be. Nevertheless, your moral rationality is clearly terrible in this case. That is an observation that the wielder of Austin's Razor can accept while denying that you have conclusive moral reasons not to pull the trigger of an empty gun and risk nothing but your own reputation.

So cases like *Three Envelopes* and *Case II* cannot help to motivate a view that would bridge the gap between acting (ir)rationally and acting in(/out of) accordance with the real reasons—or doing what one ought('nt) to do. It will not help to solve the problems of deontic significance. Of course, I do think that an appearance-relative account of *rational* action is the right view. But genuine reasons and oughts are not generally sensitive to appearances in the way that one would think if one overgeneralized from *Three Envelopes* and *Case II*. Contrary intuitions in other cases militate against this overgeneralization. Friends of this idea have been far too hasty.

Indeed, recognizing that facts like (i) and (ii) are what explain why you ought to act as you ought in some cases of factual ignorance suggests that close variants of *Three Envelopes* and *Case II* will fail to motivate a fully general appearance-relative or evidence-relative account of oughts and reasons. Suppose that in a variation on *Three Envelopes*, you ignore what you're told and form a dogmatically rigid, irrational true belief that the $1000 is in the third envelope. Here it's *false* that

> if you don't pick the first envelope, you will pick objectively randomly, and so be as objectively likely to come up empty-handed as to come up with the $1000.

You won't pick randomly. You can bring about the best outcome *with objective certainty* by sticking dogmatically with your irrational true belief. Given your foolish confidence in this (reflectively) luckily true belief, it is not as if you would be flipping a coin by choosing something other than the first envelope. Crucially, that was true in the original case, since you knew you didn't know.

Of course, you can't be praised. You are irrational. But insofar as it isn't true that you will be made objectively unlikely to pick the best outcome in this case—which it isn't, because your belief is true, and you are going to stick to it and act on it—we cannot invoke the earlier style of explanation to argue that there really is most reason for you to pick the first envelope. So, we instead can run the something like the explanation we ran for *Case I*. We can use Austin's Razor and just say that your intending to choose something other than the first envelope is right, but that

you aren't praiseworthy for doing it—and are perhaps "accusable", in a prudential sense.

Without facts like (i) and (ii) to explain how ignorance can have real deontic significance, we are left with no clear reason to think that in every case where the ultimate appearances would misleadingly recommend A-ing over all other actions, A-ing would be something genuinely deontically supported. Notice furthermore that when an explanation by claims like (i) and (ii) is available, it is not as if anything like Kolodny and Broome's feared bootstrapping is suddenly rendered acceptable. For in those cases, it isn't what the appearances indicate about the reasons that is *per se* explaining what you have most reason to do. Rather, it is the *real effects* that your heeding the appearances and acknowledging your ignorance would have on your *objective likelihood* of intentionally succeeding. But *any* plausible view should allow that objective probabilities of success and failure can generate real and not merely apparent reasons for and against acts!

The relevant objective probabilities only exist in a limited range of cases. So it can't be generally true that what we would be most rational in doing will coincide with what is genuinely deontically supported. That is only true in a limited range of cases. Often what we would be *most irrational* in doing will coincide with what is most deontically supported, because we will objectively most likely to bring about the best outcome if we stick dogmatically with our irrational beliefs and act. That is enough to get the problems of deontic significance running.

2.   *Bringing out the problems of deontic significance for epistemic rationality*

So I think there remain serious problems of deontic significance for practical rationality. My goal is not to try to address them. Indeed, I suspect they are insoluble, for reasons that will emerge in the final section of this chapter. My broader goal is instead to argue that there are parallel problems about epistemic rationality, and then try to solve them in a way that turns on what I think is a large *substantive* disanalogy between the foundations of the deontic in the practical domain and the epistemic domain. I've discussed the practical case as a way of suggesting by analogy that there ought to be similar problems for epistemic rationality. As it stands, of course, we are only in a position to give an argument from analogy for a *very generic* kind of doubt about the deontic significance of epistemic rationality. In effect, we can say:

1.   In the practical case, there is a gap between (i) correctly responding to the real reasons (or the balance of some possessed subset of them, fixed by their real weights), and (ii) complying with the requirements of rationality.

2.   If there is a gap between (i) and (ii), there will be problems of deontic significance for practical rationality. Sometimes what is required by practical rationality is the reverse of what is required by the real practical reasons (or the real balance of possessed ones, as I argued at the outset in Chapter 2). That is the wrong thing to do. The right thing to do is a function of the weights of real practical reasons.

3.   So, there are severe problems of deontic significance for practical rationality. While practical rationality may have some *other* kind of significance—say, hypological or evaluative—it is deeply unclear how it could have genuine deontic significance.

4.   The gap between (i) and (ii) *owes* to (a) the fact that while rationality turns on correctly responding to *apparent reasons*, the right thing to do turns on correctly responding to real reasons (or a possessed subset of them), and (b) the fact that there can be a gap between appearance and normative reality. Since a *false appearance* of real reasons can't *per se* generate new real reasons strong enough to outweigh or undercut the real preexisting reasons (as I've now argued), it is this gap between appearance and reality that generates the problems of deontic significance.

5.   Because the *conceptual* gap between (iii) correctly responding to the real epistemic reasons (or the real balance of some possessed subset of them) and (iv) complying with the requirements of epistemic rationality is *essentially the same* as the one between (i) and (ii), there will be severe problems of deontic significance for epistemic rationality too. So, while epistemic rationality may have some *other* kind of significance—say, hypological or evaluative—it is deeply unclear how it could have genuine deontic significance.

This argument by analogy says little about *what* the problems of deontic significance for epistemic rationality will be. It says only that we ought to expect that there will be such problems, given that the notional gap that exists between reasons and rationality in the epistemic domain is exactly the same as the notional gap that exists between them in the practical domain—something I defended in Chapter 2 in giving my unified view. What might the problems be?

2.1.   *Criteria of quality for epistemic reasons in the light of the core distinctions*

The fundamental problem is a simple and clear one once we accept the distinction between responsiveness to (possessed) epistemic reasons and epistemic rationality suggested in Chapter 2. Indeed, the problems of deontic significance for epistemic rationality are in a certain way clearer and more pressing than the problems for practical rationality, once that distinction is accepted. Why? Note first that we need a criterion of *quality* for epistemic reasons—something that will tell us when an epistemic reason is *good* or *better* than others. It is plausible that what we ought epistemically to believe will turn on the balance of epistemic reasons that is generated by this criterion. For surely we epistemically ought to believe only what the best epistemic reasons (we have) suggest. Given the connections between reasons, oughts, and justification defended in Chapter 1, this is also a point about epistemic justification. Since what we are (propositionally) justified in believing is just what we are epistemically permitted to believe, and that also turns on what the best epistemic reasons (we have) suggest, (propositional) justification turns on this too.

But surely, one might think, the criterion of quality for epistemic reasons is going to have something *objective* to do with the promotion of the fundamental epistemic goals—say, the goals of increasing true belief and decreasing false belief on a proposition by proposition basis.[161] An epistemic reason R for believing P will be *good* to the extent that believing P for R makes it *objectively more likely* that one will achieve the goals with respect to P if one believes for R, and *bad* to the extent that it *fails to make it objectively more likely* that one achieve these goals with respect P if one believes for R or *makes it objectively less likely that one will achieve these goals* with respect to P if one believes for R. Surely a truly good reason R for believing P is not one of a kind that *merely seems* to help to forward the goals with respect to P-like propositions when heeded. Seeming goodness is not goodness. Indeed, if believing for R-like reasons with respect to P-like propositions is in fact a radically unreliable policy, and *merely seems* according to total ultimate appearance to help achieve the goals with respect to P-like propositions, R-like reasons will be *bad reasons* for believing P-like propositions. What else can one say? These points feel truistic.

This uncovers a crucial, ironic reversal of Cohen's "new evil demon problem" for reliabilism, and an analogue of Kolodny's puzzle about the deontic significance of requirements of practical rationality. Given the distinction between justification and rationality, and the direct connection of the former with genuine deontic significance, the far more natural way to think about the demon-deceived subject who *merely seems by the lights of the total ultimate appearance* to be getting things

---

[161]Some readers will see a similarity between the ideas in this paragraph and the ones Goldman (1980) used in an argument against internalism. This is deliberate. As I said in the Introduction, one of the inspirations for my formulation of the problems of deontic significance for epistemic rationality is this classic paper.

right but is in fact using methods that *systematically guarantee that he will get things wrong* supports a reliabilist externalism about justification, not any kind of internalism.[162] How could one have most epistemic reason to believe in accordance with policies that are *guaranteed to fail* in point of objective fact, and so guaranteed to be *terrible* according to the real criteria? Yet *justification* as a general matter goes with good reasons and real criteria, not merely apparent reasons and mock criteria. The demon-worlder's reasons for belief simply are *bad* in her world,[163] according to natural criteria that flow from fundamental goals that almost all epistemologists accept. All the same, epistemic rationality asks her to believe for these reasons.

What we see here is structurally like Kolodny's problem for one who accepts narrow-scope interpretations of the requirements of practical rationality and regards practical rationality as deontically significant as such. Those requirements can require one to do what one really oughtn't do, merely because it appears to one that one ought to do it. Unacceptable bootstrapping ensues, if one thinks these requirements have genuine deontic significance. Normative appearance will undermine normative reality. Accepting Cohen's verdicts about demon-worlders leads to a similar result: rationality will require adopting policies that systematically guarantee failing by the lights of what are naturally viewed as fundamental epistemic obligations, fixed by the standard goals. If we accept the natural goal-based criteria, we ought to find this result unacceptable, and reject Cohen's verdicts about epistemic justification. Given a distinction between rationality and justification, as well as between justification and excuse, this is something we can do without obvious extensional difficulty. Of course, this reverses Cohen's own remarks in anticipation of a similar move to this one: he claims that we really care about rationality. As I argued in Chapter 2, these remarks simply get things backwards on quite general grounds. The demon-worlder can be excused, given how things seem with respect to total ultimate appearance. The rationality of the demon-worlder's believing what he believes is a good excuse. But excusability cannot be what matters, and pretty much by definition. The same goes for epistemic rationality too, then—at least *as such*, if it was to be the final arbiter of what we ought epistemically to do in any particular case, as he assumed.

So there is a problem of deontic significance for rationality. It is completely unclear why we ought epistemically to be epistemically rational as such. Of course, there may be *contingent* links in the *actual world*—though, as we'll see in the next few subsections, these links simply fail to hold in plenty of actually likely cases. But the same goes for practical rationality, as Kolodny recognized. The contingent links, if there are any, are simply irrelevant to the question at hand:

> One might suggest […] that by conforming to rational requirements, one is more likely to believe and do what one ought. […] [T]his reason would be instrumental. One would not comply with rational requirements for their own sake, but because doing so brought about something for which one had independent reasons. Nevertheless, it might be argued, one *always* has these reasons, so the justification is general. It does not depend on particular circumstances.
>   The problem is that it is not true, in any given case, that complying with rational requirements leads one to believe and do what one ought. […] To this, the answer is bound to come that while rationality may lead one astray in any given case, it is the best policy in the long

---

[162]This is idea is not unique to me. See Bach (1986), Engel (1992), Weatherson (2008), and Littlejohn (2009).

[163]I am unconvinced by attempts to say that we "indexicalize", so that when we are tempted to claim that the demon-worlder is justified in believing what her perceptual experiences seem to indicate, what makes that tempting is that in *our* world these experiences would be reliable indicators of *our* facts. Besides being *ad hoc* and unsupported as a semantic thesis, this simply solves no problems. It is easy enough to ask what is a good reason for belief in another world, and to get intuitions that conflict with the "actualization" verdicts. Suppose, in some distant world, the *National Inquirer* is just as reliable as the *New York Times*, or some systematic liar in our world is a fully honest man in another world. It is perfectly intuitive to say that the reports of these sources in these worlds could—and do—afford good reasons for belief in them. It would be bizarre if we couldn't say this. We are interested in the predictions of philosophical theories of good reasons for belief in other worlds. Intensional adequacy is no less important than extensional adequacy.

run. If one complies with rational requirements as a rule, then over the long run one is more likely to believe and do what one has reason to.

It is not clear why this should be so. The net result of revising my attitudes in accordance with rational requirements might be to adopt many attitudes for which I have no reason, and to abandon many attitudes for which I have. In any event, even if it is true that we have this reason to comply with rational requirements as a rule, it does not follow that we have this reason to comply with them in any particular case. *Yet the 'ought' of rationality applies in each particular case. When we say that someone 'ought rationally' to have an attitude, we are saying something about what 'ought' to happen here and now.*[164]

Kolodny's observations apply here. Indeed, they apply with greater force, as we'll see even more vividly in considering particular cases later in coming subsections. We are interested here in whether epistemic rationality as such has real (epistemic) deontic significance—whether 'S ought (epistemically) rationally to believe P' implies 'S is *ex ante* justified in believing P'. "No" is the natural answer, given criteria of quality for epistemic reasons that flow directly from goals that nearly all epistemologists endorse, and the link between reasons and justification. As we'll see, the choice of fundamental goals makes little difference as long as epistemic rationality itself isn't among them. And it is very implausible that a *fundamental* epistemic goal is to be epistemically rational. Even the internalists admit this, though they unfortunately miss the implications of the admission. Consider BonJour, who, as I've noted before, assumes that rationality and justification come to the same thing, and so who has rationality in mind in this passage:

Why should we, as cognitive beings, care whether our beliefs are epistemically justified? Why is such justification something to be sought and valued? […] The basic role of justification is that of a means to truth. […] If epistemic justification were not conducive to truth […], [it] would be irrelevant to our main cognitive goal and of dubious worth. […] Epistemic justification is […] in the final analysis only an instrumental value and not an intrinsic one.[165]

Structurally, then, it is clear that an equally well-motivated problem of deontic significance arises for epistemic rationality. Unless one rejects the view that the criterion of quality for epistemic reasons is determined by what would actually help to advance the widely accepted fundamental epistemic aims on a proposition by proposition basis, I see no way out of this problem.

In the coming chapters, this will be my radical strategy: I'll argue against the deep-seated assumption that the fundamental epistemic obligation is a teleological one, and indeed deny that teleology of any kind has a fundamental place in epistemology. Since I think we should not do something similar in the practical sphere, there is a substantive disanalogy. Right now, however, I want to be neutral, and make vivid how *once* we accept a distinction between reasons and rationality like the one widely accepted in the practical sphere, a problem of deontic significance follows from assumptions that are widely held. To my knowledge, no one except Selim Berker has tried to question the thought that the fundamental epistemic obligation is a teleological one. With a reasons/rationality distinction, a plausible way of viewing the teleological obligation is objectively: *good* epistemic reasons will be ones that *really are likely to help in the particular case.*

Now, it is worth noting that a related upshot can be reached without explicitly appealing to teleology. I flagged this point in discussing my Compromise View about the ontology of epistemic reasons in §2.3 of Chapter 1, and the move by which a defender of this view can explain the symmetry between good and bad cases of perceptual belief. Let me rehearse this.

Recall that, on the Compromise View, we drew a distinction between something's *providing* a reason

---

[164]Kolodny (2005: 543).
[165]BonJour (1985: 7−8).

and its *being* a reason. Mental states *provide* reasons. The reasons they provide are, in good cases, *facts*. This view allows us to accommodate the McDowellian idea that perception opens us up to the facts, and makes these available as reasons. The space of reasons goes well beyond the mind. This also creates the appropriate symmetry between the practical and epistemic domains. While some philosophers of practical reason continue to think that what *make* things reasons for action are mental states, it is now widely believed—largely owing to Dancy (2000, 2004)—that anyone who wants to allow that our motivating reasons can be *good* reasons will have to allow that they are worldly facts in good cases. Indeed, even defenders of the most radical desire-based views like Mark Schroeder explicitly insist that desires are just *background conditions* on reasons, and not often *themselves* reasons. On his view, while the *explanation* of why the fact that there is dancing at the part is a reason for you to go there might be your desire to dance, that hardly shows that the fact of dancing itself can't be *the reason* you have for going to the party.

The Compromise View leads to the thought that in good cases of perceptual belief, our reason for believing that there is a cup in front of us can simply be the fact that there is a cup there. As I argued before, concerns about *circularity* are not fundamental here. *Everyone* will agree that we can advert to the fact that we are in pain to justify our belief that we are in pain. Only intuitions that support indirect realism about perception, as well as skepticism about justified perceptual belief, could create strong pressure against saying exactly the same thing in the perceptual case. So the fundamental worry if there is one just isn't one about circularity. As I noted, though, the symmetry that the Compromise View creates with practical cases does have a different surprising conclusion. Your motivating reason for leaving the party could be that there isn't going to be dancing there regardless of whether you would "sadly be mistaken about that", in Dancy's words. But surely *how justified* you are in leaving—*how good* your reason is—will turn on whether or not you are sadly mistaken about that. Sure, you are equally rational either way, provided that the appearances are the same, as they will be by stipulation. That is how one accommodates the appearance of normative symmetry. But, more deeply, there is asymmetry. How, after all, could a motivating reason be *as good* if it isn't true? The same thoughts apply to the difference between the good and bad cases of perceptual belief, where the bad case is, say, the demon world. The same distinctions help to preserve the apparent normative symmetry.

So the Compromise View pressures thinking that epistemic rationality and what we have epistemic reason to believe diverge, at least in *degree*.[166] Independently of thoughts about teleological grounds for criteria of quality of epistemic reasons, there is plausibility in the general thought that one's motivating reasons cannot be *as good* if they are not true. It is also plausible that one's own motivating reasons don't change in the switch from the good to the bad case. If that's right, one's reasons are not *as good* in the two cases. Yet one is equally rational in both cases—at least if we are honest about rationality, and don't change the subject. So rationality calls for what fails to fully align with the real balance of possessed reasons, generated by their independent weights. It calls for you to do the wrong thing.

Excusably, to be sure. But, again, excusability is not our interest. So the problems may be deeper. Of course, whatever motivates rejecting goal-oriented criteria of quality for reasons might also

---

[166]Some will doubtless note that the fact that it merely seems to you that you are seeing something is still *some* reason. Now I don't think it is your *motivating* reason, given the transparency of experience: we think about the contents of our experiences, and not often the experiences *per se*, though we may be forced to appeal to them when there is nothing else. If anything, the nature of the experience explains how you get to have something *other than the experience itself* as a possible motivating reason. But the point could simply be modified: its presence could make your actual motivating reason still have some weight, even if it is a false proposition. I agree. My point need only be about *degree*. You're rational to the same degree in both cases. You're *not* forming attitudes equally well in accordance with the real weights of the reasons in both cases. Since you ought to align your attitudes with the real weights, rationality requires you to what you oughtn't.

help us to cast doubt on the thought that only motivating epistemic reasons that are true can be truly good, or that one's reasons are not *as good* in the good case and bad case. Indeed, I will argue for this conclusion in later chapters. The point for now is simply that some of the problems can be motivated without any *explicit* appeal to teleology, and it is an open and interesting question whether someone who wants to solve them will be able to undermine this further motivation if he can undermine the one grounded in teleology.

2.2.   *Illustrations*: *bootstrapping with memory*

I just offered a couple of big picture ways to see how there could be serious problems of deontic significance for epistemic rationality with the same structure as the problems in the practical sphere. But we can also bring out problems of deontic significance just by focusing on particular cases, without invoking big picture thinking about goal-oriented criteria of quality for reasons, or about the implications of accepting the right ontology of epistemic reasons. The first two cases I'll discuss were, it is worth noting, precisely the ones Alex Jackson used to undermine "Arguments from Irrationality" for seemings internalism about epistemic justification. My use of these cases is certainly informed by his excellent (2011) paper, which anyone interested in the deeper issues behind internalism/externalism disputes ought to read.

Here is one nice example discussed by Jackson that was first used in Greco (2005)'s case against internalism. Maria has a clear apparent memory that Dean Martin is Italian. She could, even after indefinite reflection, discover no apparent reason to doubt it. As it happens, the reason *why* she has this clear apparent memory is that she once acquired the belief that Dean Martin is Italian from someone that she knew to be wildly untrustworthy. She just can't remember this fact about the origin of her apparent memory. Notice that in this case it is highly plausible that

> (A)   it would be irrational for Maria to *suspend judgment* on the proposition that Dean Martin is Italian,

and that

> (B)   it would be irrational for Maria to *disbelieve* that Dean Martin is Italian.

If (A) and (B) are true, it would be irrational for Maria to do anything except believe Dean Martin is Italian, insofar as she is considering the question and looking to take a doxastic stand. Believing, disbelieving and suspending judgment are, after all, mutually exclusive alternatives, insofar as someone is considering a question and looking to take a doxastic stand. So, if we assume Mary is considering the question and looking to take a doxastic stand, it follows that:

> (C)   Maria ought rationally to believe that Dean Martin is Italian.

In his reply to Greco, Feldman (2005) used this as an argument for thinking that Maria is justified in believing Dean Martin is Italian. As Jackson noted—and as I agree—this is fallacious.

This is a case where, independently of any view about the character of the fundamental epistemic obligation and its relation to criteria of quality for epistemic reasons, we can raise a problem about the deontic significance of epistemic rationality. Intuitively, Greco was exactly right. Maria is *not* justified in believing Dean Martin is Italian. It is equally plausible to claim that if Maria appeals to the apparent memory, she is appealing to a *bad reason* for believing that Dean Martin is Italian. Bad, precisely because of its ancestry. Indeed, the two thoughts go hand in hand. If having something as your reason for belief would fail to make that belief justified, two things could be awry: (i) you could be failing to track what makes the reason sufficiently good, or (ii) the reason could simply

fail to be sufficiently good. Our case is not a (i)-type case. If all of this is right, and rationality can require us to believe for terrible epistemic reasons, that is a strike against the thought that its requirements have the kind of deontic significance one might have thought them to have. At most, the fact that Maria is rationally required to believe here is an *excuse* for something that she does *unjustifiedly*, and not a justification for an epistemic wrong.

2.3. *Illustrations: the cognitive penetrability of nondoxastic appearance*

Another illustration that has the same structure comes from reflecting on cases where the phenomenal content of perceptual experience is cognitively penetrated in a certain way—specifically, cases where things phenomenally seem a certain way *only because* you antecedently had certain beliefs or expectations. The idea of appealing to this fact to cast doubt on internalist theories of justification like Pryor's and Huemer's comes from Siegel (forthcoming), and Jackson also suggested extending the point to undermine for such theories in passing in his (2011).[167]

Consider this case. An artist has painted an array of dots on a canvas, and has called it "The Hidden". Someone looking at "The Hidden" could aspect-switch it into appearing different ways, just like with Necker's cube. Indeed, the range of possible appearances is vast: one could aspect-switch into seeing at least thirty different figures. Suppose that Dave was told a few days ago by a friend he knew to be an untrustworthy dilettante that the artist really had a dog in mind in making this painting, even though one could see almost thirty other things. Being epistemically indolent at the time, Dave unthinkingly accepted this judgment. Accepting it, let's suppose, subconsciously primed Dave so that he would see a dog when he looks at "The Hidden". Suppose that, in looking at "The Hidden" now, Dave has forgotten what he was told before by his dilettante friend. And suppose that Dave is now told by someone trustworthy that the artist had a single subject matter in mind—just not that the artist called the painting "The Hidden" because one couldn't guess it—and he is being asked to say what he thinks it is. Now,

> given the way things look to him and his further knowledge, it would be irrational for Dave to withhold from making any judgment about what subject matter the artist had in mind,

and

> it would also be irrational for Dave to disbelieve that the artist had a dog in mind.

As with the illustration in the last subsection, we can get from these claims to the conclusion that insofar as Dave is going to come to a judgment at all, he *ought rationally* to judge that the artist had a dog in mind. Intuitively, though, it doesn't follow that he would be *justified* in thinking this.

Indeed, knowing what we do about *why* things seem this way to Dave, it seems that he would be unjustified in thinking this. Given this fact about the ancestry of Dave's appearance, and the simpler fact—which Dave would have known if he had remembered what he was told—that this painting can be seen in so many different ways, this looks like a bad reason. So, as with the illustration from the last section, it looks like rationality can require us to believe for bad reasons, and to adopt unjustified beliefs. This is another a strike against the thought that its requirements have the kind of deontic significance one might have expected them to have.

2.4. *Forgotten evidence and related problems*

A related though inverted problem is illustrated by cases where one's original evidence was

---

[167]He did so in a footnote (see his (2011: n.16)), and didn't spell out the argument as completely as he did in the memory case. But it's pretty obvious that it works in the same way, as we'll see.

forgotten, but where it remains plausible that one knows—e.g., in the forgotten evidence cases that Goldman (1999) used against current time-slice internalisms about epistemic justification.

A lot of what we ordinarily claim—and certainly seem—to know and justifiedly believe rests on reasons that have faded from view. Skepticism is tempted by the thought that we don't know or justifiedly believe in these cases. But there is a real tension in these cases that externalists ignore. They evidently don't care that it would be irrational to keep believing with a similarly high degree of confidence when one is asked why one thinks what one thinks and realizes that one cannot bring to mind anything except perhaps the bare fact remembered, or the fact that one recalls—facts which will not ultimately appear to one to be available reasons in many of these cases, given the fading of ultimately apparent reasons to think that one *really is recalling something*. They also do not notice anything troubling in the counterfactual point that *if* one reflected on the question of whether one has reasons for these beliefs, it would in many cases appear that one does not, so that there are no ultimately apparent sufficient reasons for these beliefs. Nevertheless, they are surely right that we know and justifiedly believe. If it is rationality that is requiring us to give up knowledge and justified belief in these cases, that is just another reason for being skeptical that it has deontic significance as such. Notice that cases like this show that the core problem of significance is not one that arises *just* given a background view on which our epistemic goals are believing the true and not believing the false, proposition by proposition. Let the goal be knowledge. These cases show that rationality can systematically require ditching it.

Now, there is lot of psychological research on the so-called "feeling of knowing", particularly in association with the "tip of the tongue" phenomenon.[168] Jennifer Nagel (ms) has suggested that internalists about justification might help themselves to this research by claiming that even when the evidence has been lost, the feeling of knowing—which the research suggests largely to be underpinned by a *sense of familiarity*—might be an internally accessible factor that could warrant the belief that one really recalls something. She also suggests that there is a burden against externalists: were it not for the feeling of knowing, one wouldn't, she thinks, intuitively know.

Does this really help? First of all, Nagel crucially seems to neglect the point from Goldman mentioned at the end of Chapter 2. Someone might have one of these feelings of knowing in a case where that person is in fact just calling to mind a stored belief irresponsibly acquired from some obviously terrible source. If it is justification that interests us, it is doubtful that these feelings of knowing will do the work. They won't solve the problem of deontic significance.

Secondly, there is another side to Nagel's observation. In addition to feelings of *knowing*, there are also feelings of *not knowing*. These are underpinned by similar meta-cognitive mechanisms. These often explain subjects' judgments that they don't know answers to questions. Now consider a subject with remarkable first-order competence, but with a reflectively very under-confident disposition—a subject whose self-doubt manifests at the reflective level upon being questioned, but not in workaday first-order activity, where she relies unreflectively on highly competently formed stored beliefs and acts efficiently. If prompted for her reasons for thinking what she thinks, let's suppose that her meta-cognitive signals will misfire due to her misguided lack of confidence. Ultimate appearances of ignorance will overcome her case after case. These appearances will be highly unreliable by stipulation. For any propositions that she truly believes and antecedently knows, they will deliver the incorrect verdicts about whether she knows or has sufficient reasons to believe. It is hard to see how the fact that such a signal shows up in her cognitive life could be a *good reason* for abandoning her beliefs. The meta-cognitive processes at issue were, after all, supposed to be highly unreliable. She is *wrongly* under-confident at the reflective level. She was

---

[168]See Dunlosky and Bjork (2008) for some nice survey pieces on this research, and Efklides and Misailidi (2010) for some of the cutting edge research. Nagel's paper also has a terrific overview of the literature.

supposed to know at the first-order in most of these cases. Nevertheless, in the face of a sufficiently powerful meta-cognitive signal with nothing to appeal to except her stored beliefs as a stubborn rebuttal—one that will appear to her *totally arbitrary* when pressed at the second-order— it is plausible that rationality requires her to withhold belief.

So she would be irrational in continuing to believe. Does she *thereby* cease to know at the first order? Was our stipulation illegitimate? No. She can return to ordinary workaday activity and act on these beliefs after irrationally resisting these second-order challenges. When she acts on these beliefs, she does so properly. That is precisely because she still knows. At most, she ceases to know that she knows. One can, however, know without knowing one knows. The kind of internalism we are addressing is not one that embraces the clearly absurd KK principle—clearly absurd, Williamsonian intricacies aside, because it entails bad regresses, as noted in Chapter 2.

This is all, it is worth noting, just an inversion of the point made in §2.2. The point there was that rationality can require one to keep unjustified beliefs. The new point is that rationality can require one to abandon lots of knowledge, and *ipso facto* justified beliefs. The first point is one which should lead us to doubt whether Nagel is right to think that a feeling of familiarity will be able to carry the day for internalists about *justification*. And a parallel idea can be used to revive the sketch of the inverted problem that I originally pressed. Neither can be so easily averted.

2.5.   *Illustrations: self-doubt and the appearance of lost reasons*

A related problem can be gleaned separately from a not unfamiliar feature of life. The best of us are wrongly swept into self-doubt from time to time. This doubt often has a self-sufficient character: the very fact that we are self-doubting consumes cognitive resources in such a way that we run into a total block when questioned, and simply lose access to our reasons. Consider an example familiar to some of us. One spends hours intensely reflecting on a philosophical question, and comes up with several good arguments for an answer. Later, one finds oneself in a situation where one is antecedently prone to self-doubt, and when asked what one thinks and why, one draws a total blank, and is unable to bring anything to mind. From the perspective of the current time-slice, it appears that one's reasons have been lost. It may even appear unlikely that one had good reasons in the first place, since one's present incompetence may make it look as if one didn't have the intellectual resources to gain such reasons in the first place. Everything that one can summon to mind now simply looks inchoate or terrible.

It is not implausible to think that it would be epistemically irrational for one to simply stick with one's beliefs in circumstances of this kind. After all, all of one's original reasons have been rendered invisible, and one's clear incompetence on the spot rationally undercuts a generic appeal to past competence, or to the stubborn claim that there just are sufficient reasons for believing what one believes. What we see is, in effect, negative bootstrapping. Self-doubt wrongly enters the cognitive life of one whose belief was justified, renders the grounds invisible, thereby making it appear that no reasons are had, and thereby rendering the doubt rational.

But it is also hard to think that one's beliefs are hereby *unjustified,* and that one suddenly has sufficient reasons for ditching them. After all, the self-doubt was *itself unjustified* to begin with, precisely because the original reasons for belief were *good*. Once one gets out of this situation, one may rightly become frustrated. One did have good reasons! One wasn't crazy in believing what one did! How can unjustified opposition to a justified belief make it unjustified? The natural thing to think is that it simply cannot. If so, rationality can require abandoning justified beliefs— even knowledgeable ones—and to persist in doubts we oughtn't epistemically to have.

3.   *The importance of the problems of deontic significance for epistemology*

Appreciating the problems of deontic significance in the light of a reasons/rationality distinction leads to the restructuring of some orthodox ways of viewing disagreements in epistemology. One of the key places where it matters is in discussions of internalism and externalism about epistemic justification. But the implications are broader, and matter even once we've settled on a particular theory—say, an externalist theory—of epistemic justification. Before setting out to solve these problems by questioning the foundational assumptions on which they rest, it is worth pausing to reflect on the importance of this way of reorienting epistemology.

3.1.   *Reorienting internalism/externalism disputes*

Disagreements between internalists and externalists have usually been framed as disagreements about the nature of a single property—centrally, epistemic justification. It is usually thought that once we settle on a methodology for systematizing first-order intuitions—particularist or methodist—we can simply apply this methodology to cases, and then decide between accounts on the basis of their extensional adequacy, relative to whatever other theoretical virtues the methodology suggests to be key. Externalists are often particularists, and prize specific case intuitions that lead them to reject accessibility constraints on justifiers. As an example, consider the pivotal role of forgotten evidence cases as wielded by Goldman (1999). Some internalists— ones who are not skeptics—are particularists too, and just place more emphasis on *different* cases. We see this with Cohen (1984)'s emphasis on our nonfactive mental twins in demon worlds.

As I emphasized in the last chapter, once the distinction between believing rationally and believing for sufficient reasons is accepted as an instance of a broader and very well motivated distinction, this way of framing the disputes is markedly unsatisfactory. Internalists have frequently conflated intuitions about rationality with intuitions about epistemic justifiedness, often in egregiously explicit ways. It is open to externalists to simply grant *these* intuitions as being correct for rationality, to note the symmetry between the internalists' arguments and those of internalists in the literature on practical reasons (e.g., in Williams (1981)), and to take relish in how forceful Scanlon's dismissal of Williams's arguments for internalism were, given the latter's conflations. They can note that there is a general problem about the deontic significance of rationality, and argue—as I've done— that it extends to epistemology with even greater force.

So extensional adequacy can't really be the core issue. It can be granted that many internalists get the extension of *one* thing right—viz., "rational". The challenge is to motivate rejecting the criteria of quality for epistemic reasons and particular case judgments from the last few sections that suggest that epistemic rationality can't be deontically significant as such. Conceding that they get things right about rationality does not lead to some irenic dissolution of the debates. It instead brings them to a breaking point. Because epistemic justification goes with good normative reasons and so, necessarily, with genuine deontic significance, internalists are by default on the losing side. The property they care about may have a role. But it is not, without much further argument, going to be with respect to questions about what we ought to believe.

Of course, because *some* internalists also conflate hypological questions with deontic questions (as I noted in BonJour's case earlier) and love emphasizing intuitions about (ir)responsibility, they might seem to be OK with this. Perhaps they will be willing to grant that rationality is primarily relevant to questions about the hypological epistemic standings of particular subjects. But I can't imagine that once the weight of this concession is appreciated, they would all be happy about it. Some of them clearly think that it is counterintuitive to claim that the subject in the demon world is forming beliefs in the wrong way, and oughtn't to be forming beliefs as she does.[169]

---

[169]Pryor (2001: 117) is explicit: "Many philosophers share Cohen's intuition that it's possible for a brain in a vat,

3.2. *Defeat clauses in typical externalist accounts of* ultima facie *justification*

Besides forcing reconsideration of the terms of internalism/externalism disputes, recognizing the problems of deontic significance also demands much greater care than has been typical in framing clauses about defeat in standard externalist theories of *ultima facie* justification, and in using these clauses to accommodate a limited range of internalist intuitions.

Let me bring out why this is so. As I argued in Chapter 1, it is perfectly open to externalists to accept a high-level account of justification in terms of sufficient reasons. This is a second-order theory about the relations *between normative properties*, and externalists *per se* are interested in giving a first-order normative theory about the relations between the normative and the *nonnormative*. The difference is akin to the difference between a high-level account of value like Scanlon's buck-passing account on which being valuable *just is* having properties that give everyone reasons for pro-responses, and a substantive account of value like hedonism on which what makes things valuable is their being pleasurable or conducing to pleasure. They can understand the defeat of justification as a transition from having sufficient reasons to having insufficient reasons, due to the presence of new considerations. What makes their view externalist can simply be (in part) the way it understands the *quality* and *sufficiency* of reasons.

With this understood, we can also note that externalists—particularly reliabilist externalists—are in a nice position when it comes to cashing out what it takes for a reason to be *prima facie* good. For they have an extremely straightforward way of approaching questions of quality: a reason for believing that P is good only if its obtaining (*ceteris paribus*) makes a positive objective probabilistic difference to the truth of P. This claim is motivated by the simple idea that unless something *actually helps* to increase the chance that one would fulfill the fundamental epistemic obligation of believing the truth and not believing the false on a proposition by proposition basis, it couldn't be a serious consideration bearing on how to conduct one's doxastic life in that case. It would be obviously bizarre for them to accept this objective account of the **goodness** of epistemic reasons without accepting an equally objective account of the **betterness** of some epistemic reasons in comparison with others. Besides looking incoherent, this would open the theory up to the problems of deontic significance, and provide reasons to doubt whether the account could really be a correct account of justification at all.

---

if he conducts his affairs properly, to have many justified (albeit false) beliefs about his environment. And— at least in my case—this intuition survives the recognition that being *epistemically blameless* does not suffice for being justified. It doesn't seem merely to be the case that the brain in a vat can form beliefs in a way that is epistemically blameless. It also seems to be the case that he can form beliefs in a way that is epistemically *proper*, and that the beliefs he so forms would be *fully justified*—despite the fact that they are reliably false." Unfortunately, in arguing for the hardly obvious final claim that the brain in a vat's beliefs would be "*fully justified*", Pryor uses an example that process reliabilists can accommodate in weaker terms.

He considers the difference between (i) a reckless brain in a vat that believes whatever it wants, (ii) a brain in a vat that does the best it can by its lights but uses standards of reasoning it cannot tell are defective that it learned from otherwise trustworthy apparent sources—Pryor's example is statistical reasoning that fails to distinguish false negatives from false positives—and (iii) a brain in a vat who uses standards that we would deem canonically non-defective in this world. His intuition is that (i) is epistemically blameworthy, (ii) is epistemically blameless, and (iii) is better off in some way than the second. But this example does not support Pryor's strong internalist verdicts. Sophisticated process reliabilists like Goldman view *conditional reliability* as one criterion for *good inferential reasoning*: reasoning such that *if* the premise beliefs were true, the conclusion beliefs would be (objectively likely to be) true. Although Pryor is right that all brains form beliefs in ways that are "reliably false" in an *unconditional* sense, one can appeal to a difference in the conditional reliability of the inferential processes used by the second and third brains. Of course, the process reliabilist may not say that the output beliefs are *fully justified*. She will have to admit that the *input* beliefs in *non-inferential* cases cannot be justified. But it was Pyor's job to argue for that unobvious stronger claim! His example doesn't do the job. It points to a difference the process reliabilist can explain.

But if that is right, then there is going to be pressure to think that *unless* the addition of a putative consideration R* against believing that P would do something to make it objectively likely that one would *fail* in fulfilling the fundamental epistemic obligation with respect to P, then adding R* cannot make some further consideration R that is a *prima facie* good reason for believing P by theory's plausible lights *less good* or *worse*, ultimately speaking. So it can't make the *prima facie* reasons *for* believing P *less than sufficient*, ultimately speaking. In a platitudinous slogan:

> a *prima facie* bad reason against believing P cannot make a *prima facie* good for believing P a worse candidate for being an *ultima facie* good reason for believing P.

Since the criteria of quality ought to be uniform on pain of incoherence, a reason against believing P had better make an objective difference to the likelihood that one would fulfill the fundamental epistemic obligation in a given case if it is to be a real defeater. If it would be no less objectively likely given R* that one would believe a truth in believing that P on the basis of R, and R alone would make it sufficiently objectively likely that one would believe a truth in believing P, how could R* make it the case that R is not a sufficient reason for believing P? It is hard to see how one can coherently answer this question with anything other than "It couldn't".

Strikingly, however, most externalists have often *not* understood defeat in this way. This makes it unclear how they can avoid some problems of deontic significance, and how they live up to what strikes me as the best foundational motivation for their approach, which is precisely that it affords the most transparent escape route. This is notoriously true of Bergmann (2006) and Plantinga (1993)'s proper functionalist accounts. They think believed defeaters are *ipso facto* real defeaters. This claim is by itself extremely implausible, and one hardly needs to invoke the considerations I'm invoking to see why.[170] But something similar is true even on Goldman (1986)'s classic process reliabilist account. Discussing a case where someone believes her visual powers to be seriously impaired, he said: "What she believes, then, is such that *if* if were true, the beliefs in question (her visually formed beliefs) *would not* be permitted by a right rule system. Satisfaction of this condition, I now propose, is sufficient to undermine permittedness."[171]

And it is more importantly true of familiar externalist attempts to accommodate more paradigmatic internalist intuitions. Forget about beliefs, and consider the role of nondoxastic appearances. Externalists will typically take it that it's fine to allow experiences to be defeaters. To take just one example, consider Goldman (2011: 272), who is here discussing an example involving a certain Sidney who continues to believe that it is going to be sunny this afternoon:

> [H]e continues to believe an updated version of this proposition—namely, that it is sunny right now (in the middle of the afternoon)—despite the fact that he is walking in the middle of a rainstorm. Surely his current perceptual experience is a defeater for this belief […]. If we want to

---

[170]See Alston (2002a), Fumerton (2007) and Markie (2008) for complaints about Plantinga and Bergmann. It helps to think about rational requirements to see what is going wrong in their thinking. The only clear intuition behind Plantinga and Bergmann's claims is that there is something bad about believing that you have a sufficient reason against believing P and still believing P. But what's bad about *that* is that it's *incoherent*. Clearly, one could simply accept a wide-scope principle to explain why this is problematic, and that wouldn't commit one to any claims about defeat: rationality requires *not* being in this conjunction of states. We end up with bootstrapping otherwise, so that a state for which they may be no reasons, and strong contrary reasons, can give rise to rational pressure for a further state that it concerns. The bootstrapping is implausible *even without* the assumption of robust deontic significance, since it is simply implausible that an *irrational* belief about the reasons for A-ing could give rise to a rational requirement for or against A-ing. Plantinga and Bergmann leave it open that this can be true; indeed, Bergmann (2006) explicitly argues that it can be true, and his arguments are defective due to his failure to see the difference between the problem about *incoherence* that lies in the *conjunction* of states, and a problem *with one of the states*.

[171]Goldman (1986: 111).

> say what defeats Sidney's current justification for his sunniness belief, the obvious candidate is his perceptual experience.

Of course, if we just focus on (mostly) fortunate subjects like you and me, making this admission will not give rise to obvious problems of deontic significance. Our visual experiences are reliable indicators of the facts. But it easy enough to imagine subjects whose perceptual faculties are systematically defective, and in a way that they could not possibly detect. The intuition is going to remain that perceptually defective analogues of Goldman's subject who acquired the weather beliefs from nondefective sources—e.g., testimony—would be exactly as irrational if they kept the beliefs while finding themselves faced with experiences like the ones had by the subject in this case. Yet the "no defeaters" clause should not be invoked without further comment to explain what is going amiss in this kind of case. After all, if these subjects' experiences *aren't* reliable indicators of the facts, it is impossible to see how the driving standards of the theory could coherently deem these *good reasons* for abandoning otherwise reliably formed beliefs in these cases, reasons that could outweigh and make less than sufficiently good at the end of the day the *prima facie* reasons that otherwise exist in favor of believing.

I think it would be better to cautiously explain away the intuitions by appeal to requirements of rationality in these cases, and then to agree that there is simply a problem about the deontic significance of rationality that makes it unclear whether there is any deep motivation for taking intuitions in these cases to be revealing anything important about *ultima facie* justification. This point may be far-reaching. If it is the facts of reliable indicatorship that explain how a reason given by a state can be a *good* one, and not facts about, say, the state's phenomenal character or content, we're going to have to take on surprising commitments. Take a subject whose visual faculties are systematically defective and never token correct representations of her environment, and who seems to see an apple when in fact there is a dog in front of her. Suppose, moreover, that this subject *also* has a *reliable* faculty akin to BonJour's imagined clairvoyance that generates the conflicting nondoxastic appearance that there is a dog out there. *This* seeming is going to be a much better reason. So, on the whole, she will be *most* justified in believing that there is a dog out there. Barring a solution to the problems of deontic significance, which flow naturally from goal-determined criteria of quality for reasons, we will have to embrace conclusions like this.

Once such conclusions are embraced, it becomes unclear how we ought to think about familiar putative counterexamples to simple reliabilism, such as the example of Norman the clairvoyant. It seems like it should be easy enough to imagine variations on this case where the core deontic foundations that otherwise so plausibly ground process reliabilism yield the verdict that he would be justified. Given the gaps we've already accepted between justification and rationality, it also becomes rather unclear why BonJour's intuitions should be a decisive objection to the view.

Of course, this cuts in both directions. It makes very clear how counterintuitive it is to think that rationality can't be generally deontically significant, even if that conclusion follows from simple reflection on the relationship between criteria of quality for epistemic reasons and what are almost universally taken to be the fundamental epistemic duties. This is one reason why I'm going to be trying to *solve* the problem of deontic significance in the coming chapters, rather than simply leaving things like Kolodny and Broome left them in the practical case. As we'll see, the upshot will involve significant departures from core themes of the best-developed externalisms.

3.3.   *Higher-order "evidence"*

Another place where recognition of the problems of deontic significance matters is in recent discussions of the impact of "higher-order evidence". Of course, rationality as I understand it involves responding to ultimate nondoxastic appearance and not to evidence *per se*. Still, as we'll see, some of the cases that have been discussed are cases in which there are appearances *vis-à-vis*

the deontic facts which may not really be good evidence for conclusions about these facts. And indeed, what we'll see is that *unless* the problems of deontic significance can be solved, we ought to doubt whether agents truly ought to adjust their *first-order* beliefs in response to the higher-order appearances in these cases. Indeed, even if the appearances are also *good* evidence for beliefs about the first-order evidence, only a conflation of rationality in first-order belief and responsiveness to real reasons for first-order beliefs could can force one to accept some of the conclusions people have drawn in discussing higher-order evidence.

The rationale for this last claim is straightforwardly related to some earlier points. It is widely recognized that what is going on in cases where clear higher-order evidence seems to negatively impact the status of first-order beliefs is not undercutting defeat of the familiar kind. To see why, consider the following case from Christensen (2010: 187):

> (*Drugs*) "I am asked to be a subject in an experiment. Subjects are given a drug, and then asked to draw conclusions about simple logical puzzles. The drug has been shown to degrade people's performance in just this kind of task quite sharply. […] This sounds like fun, so I accept the offer, and, after sipping a coffee while reading the consent form, I tell them I'm ready to begin. Before giving me any pills, they give me a practice question:
>
>> Suppose all bulls are fierce and Ferdinand is not a fierce bull. Which of the following must be true? (a) Ferdinand is fierce; (b) Ferdinand is not fierce; (c) Ferdinand is a bull; (d) Ferdinand is not a bull.
>
> I become extremely confident that the answer is that only (d) must be true. But then I'm told that the coffee they gave me actually was laced with the drug. My confidence that the answer is "only (d)" drops dramatically."

It is clear that it would be irrational to maintain the same degree of confidence in (d) here. This is the *primary intuition* Christensen expresses in discussing these cases. Yet note that the objective probability that my belief that (d) is true *conditional on the first-order evidence and the (apparent) fact that I took the drug* is not less than the objective probability that my belief that (d) is true simply conditional on the first-order evidence. The higher-order evidence leaves everything intact. This is unlike paradigmatic undercutting defeat, where conditionalizing on the defeating information *would* destroy the *prima facie* evidential link. Schechter (2011: n.44) sums this up nicely like so:

> One way to get a grip on the contrast […] is in terms of conditional probabilities. The probability that a wall that looks red is red is presumably greater than the probability that a wall that looks red is red given that the wall is illuminated by red lights. In contrast, suppose that some premise entails some conclusion but that seeing this entailment relies on a complex bit of reasoning. The probability that the conclusion is true given that the premise is true is no greater than the probability that the conclusion is true given that the premise is true and given that I'm unreliable in the relevant kind of reasoning.

The question that needs to be addressed is whether the fact that there would be irrationality in these cases if one failed to respond to the higher-order evidence has any bearing on what belief-forming rules the subjects *ought* to use, and whether their reasons for believing the first-order claims are affected at all. Christensen and Schechter do not see this question. But it is a real one.

One thing that is worth recognizing is that the intuitions of irrationality to which Christensen and Schechter appeal remain regardless of whether the subjects in their cases really *are* defective in the ways the appearances suggest. Suppose that, in a slight variation on *Drugs*, it turned out that the experimenters were lying, and that they didn't drug the coffee. Indeed, suppose they've never done this. This has no effect at all on my intuition that it would be irrational for the subject to maintain the same degree of confidence in (d). *That* intuition remains for me.

When this is recognized, the problems of deontic significance kick in. Given that reasoning exactly as the subject reasons in *Drugs* is the *right* way to reason, and yet that it would be irrational to reason in this way, how could the fact of irrationality in disregarding the higher-order evidence have any bearing *per se* on whether or not that's what the subject *ought* to do? As it stands, it is open to one who is already willing to embrace the conclusion that rationality is deontically impotent to say that it *doesn't*, and say that *at most* the fact of irrationality would have implications for whether the subject could be *praised for doing the right thing*. Just as there can be excusable wrongdoing, as we've seen, there can be praiseless rightdoing, and even—if we accept Zimmerman's category of "accuses"—blameworthy rightdoing.

That is, after all, exactly analogous to the claim that many make in the practical case. People writing on "higher-order evidence" have an obligation to defend the deontic significance of rationality. This is crucial, given that they appeal to the *irrationality* of failing to respond *in one's first-order doxastic conduct* to clear higher-order appearances about what epistemic reasons for first-order beliefs are had to motivate rejecting otherwise extremely plausible claims about epistemic justification. Schechter, for instance, takes the irrationality of failing to heed one's own fallibility after engaging in a long chain of deductive inferences to cast doubt on *single-premise closure for justification*. This is not a principle to be lightly rejected! It is worth considering an alternative take on the intuitive data suggested by a pessimistic view about rationality's deontic significance.

4.   *The options and a preview of coming attractions*

How should we respond to the problems of deontic significance for epistemic rationality?

We could simply embrace the conclusion that epistemic rationality has no deontic significance *per se*. Instead, it always has a different kind of significance—say, hypological. On this view, while *we* are always epistemically blameworthy for epistemic irrationality, and are always epistemically excusable for mistakes as long as they are not accompanied by irrationality on our behalf, these things just come apart from questions about *what* ought to be believed, *what* we have most reason to believe, and *what* beliefs are held rightly. There is just a deep act/agent distinction of precisely the kind that people have already accepted in the ethics literature. We already knew that there is excusable wrongdoing and (perhaps) "accusable" rightdoing: the *agent* may be flawed for some choice without the *choice's* being problematic, and the *choice* may be problematic without the *agent's* being flawed for choosing it. This is, indeed, a move that I once thought advisable: the problems of deontic significance for epistemic rationality are insoluble, and we can explain away the intuitions by endorsing what I called the "strong separability thesis" at the end of Chapter 1.

If there were no other defensible options, this is still what I would recommend. Of course, as we've seen, this line would have to pushed very far to explain everything. The farther we have to push it, the less plausible it becomes. Partly for this reason, I think it's worth seeing whether there is another option. What we would need to do is to question the fundamental assumptions on which the theoretical argument for pessimism about the problems of deontic significance for epistemic rationality turned. The most fundamental of these assumptions are

> the assumption that the core epistemic duty is a teleological one—increasing epistemic goods like true belief or knowledge, and decreasing epistemic bads like false belief or ignorance, most plausibly on a proposition by proposition basis,

which is supposed to seem plausible *because* of

> the further assumption that what it is to be epistemically good is to be suitably connected in instrumental terms to certain core epistemically desirable states like true belief and false belief

(or knowledge and ignorance).

Why are these assumptions needed to create problems of deontic significance in epistemology?

As we saw earlier, these problems get off the ground in the most general way by means of the thought that we need *criteria of quality* for epistemic reasons, and the further thought that these criteria of quality are going to be given by the extent to which the presence of these reasons would be objectively relevant to satisfaction of the core epistemic duties of believing the truth and not believing the false on a proposition by proposition basis. An even more fundamental thought was this: how could it be that we ought epistemically to form or revise certain first-order attitudes if doing this would *systematically lead to fundamentally epistemically worse states of affairs across particular cases*? In formulating the puzzle, I assumed for the sake of argument that this rhetorical question was forceful, and that the constraints on criteria of quality for reasons were right. These are, after all, *prima facie* plausible thoughts that flow from doctrines that many accept.

But these assumptions are not so innocent. Indeed, I think they're mistaken, and rest on a tempting misconception of (i) the structure of epistemic value, and of (ii) the relationship between epistemic value and the deontic in epistemology. Crucially, I do not think that the analogous claims in the practical case are wrong. To me, teleology is close to correct about both the nature of value, and the relationship between value and the deontic in the practical sphere. I am seduced by (agent-neutral) consequentialism and the teleological axiology on which it rests. This is part of why I was tempted by the first response to the problems of deontic significance: it seemed obvious that the same fundamental *substantive* views should be good for epistemology. That, however, is no longer a temptation of mine. The assumption that analogues of these *substantive* claims will be equally good for epistemology is wrong, I now think. My goal in the coming chapters will be to show how wrong the assumption is and, by doing that, to show how epistemic rationality has genuine deontic significance, and how internalist constraints on the deontic can be defended even if wholesale internalism is false for other reasons.

But why do I assume that rejecting the teleological presuppositions of the argument will help in a *positive* way, and *vindicate* the deontic significance of epistemic rationality? Of course, simply rejecting them would do nothing except defeat the arguments for pessimism. But one cannot responsibly dismiss these arguments without offering an alternative. The alternative need not involve abandoning the thought that *truth* should have a fundamental place in epistemic axiology. It could amount to adopting a different stance on *what relation* to truth most fundamentally explains its place in our epistemic axiology. So far, we've been operating under the extremely common assumption that the fundamental relation is the teleological one of *promotion*. Truth-*conducivity* is supposed to be what matters, as well as falsehood-*obstructivity*. We can sum this up as

> **the teleological conception of epistemic value**, according to which what it is for true belief (or knowledge, or whatever) to be an epistemic value is for it to be the case that we ought epistemically to *produce it*, and what it is for false belief (or ignorance, or whatever) to be an epistemic disvalue is for it to be the case that we ought epistemically to *eliminate it*.

This is an instance of a broader teleological claim about the structure of value that Scanlon nicely summed up in the words, "To be (intrinsically) valuable […] is to be 'to be promoted'."[172] If we reject this claim, it is open to us to hold that there are *many* different relations to truth, understood as a property of doxastic representations, which might ground its epistemic value. Conducivity (or obstructivity) needn't be the only one that matters. A different relation of particular

---

[172]Scanlon (1998: 80). By "promoted", Scanlon means what I intend by "produce". "Produce" is really the better term. Intuitively, one can, I take it, promote art by getting people to appreciate it, and without making more of it. And Scanlon certainly did not have this broader sense of promotion in mind.

importance is one associated with the Kantian tradition: *respect*. What I want to flag now is that *if* we replace the teleological conception of the epistemic value of truth with a conception grounded in respect, we can illuminate why pessimism about the deontic significance of epistemic rationality fails. I'll explain this briefly now and save the details for later.

First, a few words are in order about the relevant notion of respect. What we need is what Stephen Darwall calls "*recognition respect*" rather than a different thing he terms "*appraisal respect*":

> There is a kind of respect which can have any of a number of different sorts of things as its object and which consists, most generally, *in a disposition to weigh appropriately in one's deliberations some feature of the thing in question and to act accordingly*. The law, someone's feelings, and social institutions with their positions and roles are examples of things which can be the object of this sort of respect. Since this kind of respect consists in *giving appropriate consideration or recognition to some feature of its object in deliberating about what to do*, I shall call it *recognition respect*.
>
> There is another attitude which differs importantly [...] which we likewise refer to by the term 'respect'. Unlike recognition respect, its exclusive objects are persons or features which are held to manifest their excellence as persons or as engaged in some specific pursuit. For example, one may have respect for someone's integrity, for someone's good qualities on the whole, or for someone as a musician. Such respect, then, consists in an attitude of positive appraisal of that person either as a person or as engaged in some particular pursuit. [...] Because this sort of respect consists in a positive appraisal of a person, or his qualities, I shall call it *appraisal respect*.[173]

What is crucial to recognize about recognition respect is that it imposes a constraint on a subject's *deliberations* about what to do that is sensitive to her *perspective* on the facts. To see what the constraint might involve, it is useful to consider an example from the practical case:

> *Ingenious Marital Reconnaissance*. Jane worries that her husband Dale would cheat on her if he had the opportunity. She decides to test this hypothesis in a clever way. Using her extraordinary costuming skills, she manages to dress up to look like a different woman on whom she suspects Dale would have an instant crush. Disguised, she starts regularly showing up outside his workplace to flirt with him. Dale believes on this misleading but compelling evidence that he is interacting with a different woman. He flirts with her in turn. Indeed, he makes plans to have a romantic evening with her when he next thinks that his wife is out of town, and not to tell his wife a word about it. That is what he thinks he is doing next as he goes through with his plans. He knows that his wife would be outraged if she found out. But he doesn't care, and shows his romantic interest to a great time.

Dale plausibly lacks the kind of respect he really ought to have for his wife. Why? He does not *actually cheat* on her, after all. He does nice things for her! She *is* the woman in disguise. His failure of respect cannot be located in any of his acts, understood as external occurrences involving other people. It is rather located in the relationship between his *motives* and *how things appear to him*—between various elements of his *perspective* on what he is doing. It is because it appears to him that he that violates the terms of his relationship with his wife, and because this has no constraining influence on his practical reasoning, that he fails to have sufficient respect.

This case shows how there could be a collapse of deontic facts into hypological facts, if we think it is intuitive that Dale ought to have respect for his wife, and that this is, indeed, a key to the survival of the relationship they have. Recall that, in the practical sphere, one usually draws the deontic-hypological distinction by claiming that while *how the reasons appear from someone's perspective and the character of her deliberation* may affect what she would be *criticizable or creditworthy* for doing, only her changes to the world's state affect whether she did what she ought to do. If it appears that giving a certain liquid to a person would kill her when it will actually save her life, and she

---

[173]Darwall (1977: 38–9); some italics mine.

drinks it at your duplicitous behest, there remains a clear sense in which *what* you do is right. You save her life, after all! But you are still plausibly open to criticism in acting, because you acted in a way that was wrong by your lights. So, the line goes, these evaluations are orthogonal. Rightness turns on the facts, whereas criticism of the agent turns on her perspective on the facts. Crucially, however, *they can't be orthogonal when what you ought to do is give recognition respect to something. Failing by your own lights entails failing objectively in this case.* Whether you count as having recognition respect turns not (just) on your effects on the world, but far more crucially on your *perspective* on them and the relation of your *deliberation* to it. Norms of respect can't, I suggest, be blamelessly violated, or conformed with in ways that don't warrant some minimal positive appraisal.

While I am extremely doubtful that the most important value-grounding response in the practical sphere is recognition respect—though it is, I think, key for understanding some special obligations like the one in *Ingenious Marital Reconnaissance*—I think there is great promise in exploiting this response to explain epistemic value. Let's just suppose for the sake of argument that it is, so that we can replace the teleological conception of epistemic value with

> **the (veritist) Kantian conception**, on which what it is for truth (understood as a property of doxastic representations) to be an epistemic value is for it to be the case that we ought epistemically to give it respect in our theoretical reasoning about answers to questions— specifically, by giving the rule *believe P only if P* full recognition respect in each case in theoretical deliberation directed at answering the question whether P.[174]

Together with the Kantian conception of epistemic value, I'd also recommend

> **a deontological conception of the truth-respecting norm**, according to which, in considering whether to believe that P, what matters is whether one has sufficient respect for the truth about the question of whether P.[175]

I'll call the conjunction of these claims **strong epistemic Kantianism**. Because strong epistemic Kantianism purports to specify the fundamental epistemic norm, and indeed the norm that spells out what it is for truth to be epistemically valuable, it offers us a way of giving a very different criterion of quality for epistemic reasons than the one that flowed from the earlier teleological assumptions. The hypothesis I'll be exploring later is that this criterion will directly entail that requirements of epistemic rationality are genuinely deontic.

Here is the thought. Satisfying requirements of epistemic rationality *just is* a necessary condition for having full respect for truth. How, after all, could an epistemically irrational subject be perfectly effective in giving the norm to believe that P only if P full weight in her deliberations? An epistemically irrational subject, on my theory from Chapter 2, will either be clearly incoherent or fail to heed apparent likelihoods of truth. It is impossible to see how one could be irrational in either of these ways while retaining full recognition respect for truth. But if that's right, we can

---

[174]Mark well that I say "*in every case*" and "*fully*". There is some plausibility in the functionalist thought that a state couldn't count as a belief if the person who is in the state weren't inclined in some way to see to it that it is true rather than false. Still, we obviously want to allow for lapses: wishful thinking, lazy reasoning, and so on, are surely not impossible. To be *inclined* in the constitutively relevant sense is not to be *fully inclined in every case*; it is certainly *not* true that one can't have a certain belief simply because, in some specific case, one shirks one's duty to fully give a damn about ensuring the truth of what would be this belief's content. For this reason, the demand that on the Kantian theory exhausts the value of truth is not a constraint that any believer automatically satisfies as a matter of constitutive necessity, though the source of the stronger demand surely has *something* to do with the constitutive aim of belief.

[175]This contrasts with a consequentialist picture of the respect-based norm, on which what really matters is just maximizing instances of norm-compliance and minimizing instances of norm-violation. This is implausible for reasons I'll bring out later in discussing Selim Berker's opposition to epistemic teleology.

simply argue as follows:

> 1.   For any S, S ought to have recognition respect for the rule *believe P only if P*. This is simply what it is to have the kind of respect for truth we ought to have.
> 2.   Complying with requirements of epistemic rationality is the constitutive means for respecting the truth in this sense.
> 3.   If S ought to A, and B-ing is the constitutive means for A-ing, then S ought to B.
> 4.   So, S ought epistemically to comply with the requirements of epistemic rationality.

If all practical values were to be understood in non-teleological, respect-based terms, we could give a simple argument for the robust deontic significance of practical rationality by generalizing our observations about what respect requires in cases like *Ingenious Marital Reconnaissance* and then giving a parallel argument to the (1–4) argument. Since I do not think anything like this is plausible for the practical sphere, I remain doubtful about the genuine deontic significance of requirements of practical rationality. In general, they are relevant only to hypological questions, not deontic ones. But I think the Kantian conception of epistemic value and the picture of the foundations of epistemic normativity it affords will give us the way out.

Such, at any rate, is a project for ensuing chapters. Whether or not this project succeeds, the implications for epistemology will be extremely important, as I've argued in this chapter. Some of the most interesting questions in epistemology—e.g., the questions more fundamentally at stake in internalism/externalism disputes—turn precisely on whether or not we ought to be pessimists about the normativity of epistemic rationality, and hence on our views about the criteria of quality for epistemic reasons, and hence, I say, on our views about epistemic teleology.

IV

WHAT EPISTEMIC VALUE CAN'T BE
IF IT CAN FIX THE DEONTIC FACTS

The idea that to be good is simply to be "to be promoted" can seem an extremely natural, even inescapable one. It is plausible to think that, as Shelly Kagan insists, the good simply is that which we have reason to promote. But although there are many cases in which this

is true [...], it becomes quite implausible to hold that all of our thinking about value can be cast in this form.

—T. M. Scanlon[176]

1.   Introductory Remarks

1.1.   *The dialectical space*

There are, as we've seen, two forms of epistemic teleology. What I've called the *strong* form can be seen as giving an objective goal-based criterion of quality for epistemic reasons. Being a bit more precise than before, a minimalistic formulation of this view would be:

> **Strong Epistemic Teleology** (*Reasons Quality Formulation*):
>
> R is a *prima facie* **good epistemic reason** for believing P *only if*, setting other epistemic reasons bearing on whether to believe P aside, believing P for R makes it objectively (more) likely that one will fulfill the **core epistemic goals** with respect to P.

The Strong Epistemic Teleologist *could* be seen as making this claim by making a more schematic claim about the relationship between the epistemically right and the epistemically good: roughly, facts about what is epistemically right are fully fixed by facts about what *promotes* the epistemically good, where the teleologist presupposes as a background constraint that the good *just is* that which is "to be promoted" or "to be brought about". Ultimate goods that are "to be promoted" or "to be brought about" are usefully conceived of as *goals,* so this idea naturally motivates the formulation of Strong Epistemic Teleology I've given. And the resulting overall picture would be closer to standard formulations of teleology in ethics after Rawls (1971).

But I prefer to view the Strong Epistemic Teleologist as putting the Reasons Quality Formulation first, partly because this is closer in spirit to what people have actually done in the epistemology literature. The view can still be usefully called teleological. As I'll argue shortly in discussing Selim Berker's case against one kind of strong epistemic teleology, it is important to avoid saddling the epistemologists who can be reasonably called "teleologists" with views they would not accept. The analogy with teleological theories of reasons for *action* can be misleading. The formulation I'm giving is preferable, as I'll argue shortly, because it sidesteps the misleading aspects of this analogy, and thus seats the people who should be called "teleologists" in epistemology with a view they either *do* accept or *ought rationally* to accept, given their theories. Indeed, the Reasons Quality Formulation is entailed or at least suggested by the conjunction of some standard externalist accounts of *prima facie* justification (e.g., indicator reliabilism and process reliabilism) with the high-level account of epistemic justification in terms of epistemic reasons I defended in the first chapter and some substantive assumptions about what the core epistemic goals are. Among views still being defended, those externalist views are most deserving of placement under the heading of "epistemic teleology".[177] If the resulting view is not precisely structurally analogous to the claims

---

[176]Scanlon (1998: 87).

[177]Why not put some *internalist* views in the same camp? Don't many internalist theories entail or at least suggest that an epistemic reason for believing P will be *prima facie* good to the extent that, setting other reasons aside, believing for that reason makes it more likely in an ultimate-evidence-relative or appearance-relative or belief-relative sense that one will fulfill the core epistemic goals with respect to P? And isn't that kind of view worth calling a form of epistemic teleology—just a *subjective* form? Well, many internalist theories may **extensionally coincide** with a subjective version of Strong Epistemic Teleology. But I think most internalists should be seen as *deriving* these extensional implications from deeper claims that are not teleological. The only obvious exceptions are Richard Foley and Keith Lehrer (see e.g. Lehrer (1974)). But their views are far from paradigmatic. Most contemporary internalists do not accept views like theirs.

that teleologists in ethics would want to make about the criteria of quality for reasons for *action*, that is going to owe to either some general structural disanalogies between reasons for *action* and reasons for *attitudes* that teleologists should accept, or to some high-level constraints on epistemic reasons that are simply antecedent to first-order theories like Strong Epistemic Teleology as I understand it.

I'll return to this momentarily. For now, note that on the assumption that the core epistemic goals are *believing the truth* and *not believing the false*, we can collapse the earlier formulation into:

> **Veritist Strong Epistemic Teleology** (*Reasons Quality Formulation*):
>
> R is a *prima facie* good epistemic reason for believing P *only if*, setting other epistemic reasons bearing on whether to believe P aside, believing P for R makes it objectively (more) likely that one will believe the truth (not believe the false) with respect to P.

I'll focus on this version of the view for simplicity's sake, though I will turn toward the end of the chapter to explain why switching to a different picture of the core epistemic goals does not help. It is also worth stressing that the objective probability constraint in this highly schematic view could be fleshed out by different specific views in different ways. The process reliabilist focuses on the reliability of the process of forming beliefs in P-like propositions for R-like reasons. The facts about reliability of this process will fix the relevant objective probabilities needed to understand the view's implications. The indicator reliabilist, on the other hand, focuses on the reliability of (true) R-like considerations in indicating P-like facts. It will not be necessary for me to target either view singly. I'll criticize only implications of both views.

It is important to distinguish Strong Epistemic Teleology from a further form of teleology:

> **Weak Epistemic Teleology**: What it is to be epistemically good is to be a goal that we ought epistemically to fulfill or to be conducive (constitutively or causally) to goals that we ought

---

As we'll see, an epistemic Kantian view on which our fundamental epistemic duty is respect for the truth makes all the extensional predictions of a plausible appearance-relative form of "subjective" Strong Epistemic Teleology. The resulting view is still not worth calling teleological. This is because it is built on more fundamental, non-teleological premises. And I think most internalists really are best seen as making such deeper assumptions. At any rate, I think they should make such deeper assumptions to avoid the criticism that they conflate epistemic rationality with justification and can't have a correct view about epistemic *justification* owing to the problems of deontic significance discussed in the last chapter.

This is, it's worth mentioning, a major source of dissatisfaction I have with Selim Berker's picture of the dialectical space. He thinks the majority of epistemologists—internalists and externalists alike—are epistemic teleologists. He says that the internalists are (often) just "subjective" teleologists, whereas externalists are (often) "objective" teleologists. This admittedly creates nice symmetry with the ethical case. Consequentialists, after all, often split into subjective and objective camps. But it misrepresents the deeper commitments of the people on the internalist side of the fence. These people do not—and indeed could not, given their need to solve the problems of deontic significance—accept "subjective teleology" as a *bedrock* claim. They may accept the *extensional implications* of subjective teleology. But they should (and do, I believe) derive these conclusions from more fundamental, non-teleological premises. If so, their views aren't worth calling teleological. Having been taught (as an undergraduate) by two dyed-in-the-wool internalists, and having myself been an internalist for a time, I'm confident they would be happy with this characterization. Foley and Lehrer are unusual cases. Many internalists will distance themselves from them.

The core point here is one about *priority*. One can see starkly what I have in mind by by analogy. Suppose for the sake of argument that Derek Parfit is right that rule consequentialism, Kantian ethics, and contractualism make all the same predictions. If Parfit is right, is Kantian ethics really just a form of rule consequentialism? No, because the fundamental, driving ideas are not consequentialist. So, from the fact that a view has all the same extensional implications as rule consequentialism, it hardly follows that the view must itself be a form of consequentialism. This is a reason why Parfit's "convergence" claim could not, if true, be the end of ethical theorizing. There would still be disputes to be had about, say, whether the Kantian's fundamental ideas are the right ones. The same point applies here with mainstream internalists.

epistemically to fulfill. Similarly, what it is to be epistemically bad is to be something that we ought epistemically to avoid bringing about, or to be conducive (constitutively or causally) to what we ought epistemically to avoid bringing about.

This claim is widely accepted or at least widely presupposed. It is presupposed most vividly by people who have, at least at certain points, been Strong Epistemic Teleologists. As I noted in Chapter 1, one sees it assumed without comment in slides like these from Alston (2005: 29):

> We evaluate something epistemically when we judge it to be more or less good or bad from the epistemic point of view, that is, *for the attainment of epistemic purposes.*
>
> [T]he evaluative aspect of epistemology involves an attempt to identify ways in which the conduct and products of our cognitive activities can be better or worse *vis-à-vis the goals of cognition.*
>
> (Italics mine.)

Weak teleology is contained in the shift from talk of *good* and *bad* to the talk of *purposes* and *goals*. This shift is certainly not local to Alston. One sees it all over the place in contemporary epistemology. Many slide between talk of epistemic goals and talk of epistemic values.

If these slides were acceptable, it would help Strong Epistemic Teleology a lot, I think. Here is one simple argument to this effect:

> *The Argument from Correctly Responding to Epistemic Value*
>
> 1.   (**Value-Driven Constraint on the Criteria of Quality for Reasons**:) R is a *prima facie* good epistemic reason to believe P *only if*, setting other reasons bearing on whether to believe P aside, believing P for R makes it objectively (more) likely that one will correctly respond to intrinsic epistemic value simply in believing P.
>
> 2.   If Weak Epistemic Teleology is true, correctly responding to epistemic value *just is* doing what would promote it as a goal (ultimate or derivative).
>
> 3.   If (1), (2) and Weak Epistemic Teleology hold, Strong Epistemic Teleology is true.
>
> 4.   So, since Weak Epistemic Teleology is true, Strong Epistemic Teleology is true.

I would encourage us to accept the Value-Driven Constraint. Indeed, I myself will rely on this constraint later on in arguing for a *non-teleological alternative* to Strong Epistemic Teleology. Given this constraint, it is easy to derive Strong Epistemic Teleology from Weak Epistemic Teleology.

What this suggests is that we can undermine Strong Epistemic Teleology if we can argue against Weak Epistemic Teleology. My strategy in this chapter will be to pursue this hypothesis by focusing most critically on Weak Epistemic Teleology.[178] Accordingly, I'll give several arguments against Weak Epistemic Teleology. These arguments will make plausible further claims that can be used to argue *directly* against Strong Epistemic Teleology. Indeed, these arguments would undermine other formulations of Strong Epistemic Teleology than the one I started with. Even if Strong Epistemic Teleology can be *prima facie* supported on other grounds than the Argument from Correctly Responding to Epistemic Value, it would fail anyway precisely on account of the reasons why Weak Epistemic Teleology fails. As we'll see, some of these arguments lend credibility to an alternative picture—what I've called *epistemic Kantianism*—that can be used to explain the deontic significance of epistemic rationality. I'll explore this in the next two chapters.

---

[178]In inspiration, I am indebted to Scanlon. In Chapter 2 of *What We Owe To Each Other*, Scanlon's method for arguing against strong teleology in the practical domain is to argue against the picture of value it presupposes, and to show that it loses its appeal once we see why that picture is false.

1.2.   *Why Berker hasn't refuted epistemic teleology*

Some informed readers may ask: "Why bother? Hasn't Selim Berker refuted Strong Epistemic Teleology in "Epistemic Teleology and the Separateness of Propositions"?" My response is that Berker has not refuted Strong Epistemic Teleology. Let me briefly explain why I think this, partly because it may clarify how I am understanding the dialectical terrain.

Berker starts with the assumption that Strong Epistemic Teleology is best formulated in a way that does not make our epistemic duties *vis-à-vis* forming a doxastic attitude with respect to P depend *only* on comparisons between the degrees of epistemic value and disvalue (*probably*) <u>inhering just in the doxastic options with respect to P themselves</u>. This is admittedly a natural thought if we are trying to erect direct structural analogies with teleological accounts of the criteria of quality for reasons for <u>*action*</u>. Strong teleology about reasons for action cannot be the view that whether there is most reason for you to A turns on the comparisons between the degrees of value *inhering in A-ing itself* and the alternatives to A-ing *themselves*. That is absurd. After all, the values of the causal consequences of A-ing and of the alternatives to A-ing will matter on any sensible view of reasons for action, not just pure consequentialist views. Yet the value <u>*inhering in A-ing itself*</u> and in the *alternatives themselves* will not turn on causal consequences.

This natural assumption leads Berker to consider at the outset a stronger view I'd put as follows

> **Strongest Epistemic Teleology**: There is most epistemic reason to believe that P iff to believe that P would (likely) be to form the epistemically best doxastic attitude to P, where the epistemically best doxastic attitude to P is one that results (constitutively/causally) in the greatest promotion of core epistemic goals and greatest avoidance of core epistemic anti-goals (e.g., false belief) among available doxastic attitudes to P.[179]

Strongest Epistemic Teleology is certainly absurd. Berker illustrates its absurdity with cases inspired by Firth (1981) and Fumerton (2001). A nice if unrealistic case from Fumerton is

> *The Demon Case*: Suppose […] that belief is under [my] voluntary control and that I know that there is an all-powerful being who will immediately cause me to believe massive falsehood *now* unless I accept the […] conclusion that there are unicorns. It would seem that to accomplish the goal of believing what is true and avoiding belief in what is false *now*, I must again adopt an epistemically irrational [and unjustified] belief.[180]

This case helps to illustrate that the problem doesn't stem from Strongest Epistemic Teleology's failure to restrict its claim about the fulfillment of goals to the time of doxastic decision. I agree with Berker that cases like this refute Strongest Epistemic Teleology and many relatives of it. Similar examples will refute any version that fails to restrict the comparisons of epistemic value that matter for determining whether to believe P to the comparisons between the epistemic value (*probably*) <u>inhering in the doxastic options <u>with respect to P</u></u>. Someone starting with Berker's picture of the dialectical space may regard restrictions of this kind to be unmotivated. What would reasonably tempt this verdict is the structural analogy with teleological theories of reasons for action noted earlier. No strong teleologist about reasons for action will say that reasons for A-ing are

---

[179]It is worth stressing that this is terminologically different from Berker's formulation. He doesn't formulate the view as a view about epistemic reasons, but rather as a view about epistemic oughts and justification. He also builds his initial formulation of Strong Epistemic Teleology out of several subtheories—a theory of contributory final epistemic value, a theory of overall epistemic value, and a deontic theory. But these differences are immaterial. My formulation differs only to streamline the discussion, and render my own use of terminology consistent. All the points I will be making apply to his explicit formulations.

[180]I lift this *verbatim* from Fumerton (2001: 55).

determined just by comparisons to comparisons between the degrees of value that *inhere* in A-ing itself and the alternatives to A-ing themselves. This is both *ad hoc* and implausible.

But one might wonder whether there is a structural disanalogy between this case and our case that would make the restriction principled. I think there is. It is one that has nothing to do with views about the foundational criteria of quality for reasons or the status of any first-order theory of reasons for attitudes, but rather with general independent constraints on the *form* that all accounts of reasons for certain attitudes such as intention, belief, envy, admiration, contempt, and so on, can take. To bring out what I have in mind, I will start with a revealing analogy.

It is not at all arbitrary for direct consequentialists about reasons for *action* to claim that reasons for A-ing completely determine which considerations bearing on *intending to A* fix the fittingness or correctness of intending to A. This claim implies that the consequences of *merely intending to A* are irrelevant to whether that intention is correct. Suppose someone would give you millions *merely for intending* to eat sickening filth, regardless of whether you actually go ahead and eat it. This is not the kind of consideration that can render an intention to eat sickening filth fitting or correct. As Derek Parfit would say, it does make it fitting and correct *to cause yourself* to intend to eat sickening filth.[181] But it is not a reason *for the intention*. Direct consequentialists of the familiar sort can accommodate this intuition by accepting the structural principle I mention.

What we see here is just an example of a much broader distinction between "right-kind" and "wrong-kind" reasons for attitudes.[182] Mark Schroeder gives some nice examples here:

> Imagine financial rewards for admiring the despicable or lauding the mediocre, or invent an evil demon with plans to punish your family if you are not afraid of Mickey Mouse or if you aren't amused by *Schindler's List*. The scenarios needn't even be unrealistic; as D'Arms and Jacobson note, there are real-world jokes that are funny even though it is wrong to be amused by them, and true-to-life scenarios in which by far the smartest course of action is to admire someone lacking in admirableness, or to withhold envy from the enviable.[183]

There is a general distinction among reasons for attitudes revealed by these kinds of cases. Only certain considerations are even possibly relevant to whether an *attitude* is fitting or correct as opposed to an *attitude-causing act*. No wise consequentialist of any familiar kind should be uncomfortable with this as a point of partisan orientation. It is a distinction that is antecedent to substantive first-order normative theorizing of the sort she is doing. It places a constraint on the *form* that first-order theories of the justification or permissibility of *attitudes* can take, as opposed to theories about the justification of *acts of causing ourselves to have attitudes*. Owing to Kavka's famous Toxin Puzzle, it is widely known that intention is no exception. The thing for practical teleologists to do if they want to give a plausible story about reasons for intention is to accept this distinction and construct their account of reasons for intention in a way that honors it.

The structural principle noted earlier was intended to illustrate how this might be done. There is nothing unprincipled about this move. Drawing the distinction between right-kind and wrong-kind reasons for attitudes is not itself a retreat from direct consequentialism of the half-plausible kind. The right-kind/wrong-kind distinction does not reflect some *deontological* restriction. It is

---

[181]See Appendix A in Parfit (2011 v.1). This idea is not unique to Parfit. One sees also it in Gibbard (1990: 37) and Skorupski (2007: 9–12). Closely related ideas are found in Hieronymi (2005) and Piller (2006).

[182]The broad distinction at issue is one that transcends local counterexamples to some buck-passing accounts of value, as Schroeder (2010) rightly notes. I would, however, disagree with him that providing an account of this more general distinction will by itself get buck-passers off the hook. For I think there is simply a different, deeper "wrong kind of reasons" problem for buck-passers. I argue for this in Sylvan (Ms₃).

[183]Schroeder (2010: 27).

not a distinction that is part of first-order normative theory at all. It simply constrains the form all first-order normative theories about attitudes can take—at least if these are not secretly going to be about whether we ought *to cause ourselves* to have the attitudes, which would be an *act*. The resulting consequentialist account of reasons for intention is still interestingly consequentialist, precisely on account of the structural principle linking reasons for action to reasons for intention and the direct consequentialist account of reasons for action.

I say that a Strong Epistemic Teleologist can accept *dialectically parallel* structural restrictions that avoid all of Berker's "separateness" counterexamples to Strongest Epistemic Teleology. To see this, let's consider one widely accepted theory of the distinction between right-kind and wrong-kind reasons for belief. This theory is an instance of a broader theory according to which all right-kind reasons for commitment-involving attitudes such as intention, belief, etc., are "object-given", as Parfit says: they are reasons that bear in some way on the objects (=contents) of these very attitudes. A natural version of this theory for belief says that any factor that affects whether we ought to believe has to be a consideration that bears (logically or probabilistically) on the truth of the content of the belief. All of Berker's putative counterexamples in "Epistemic Teleology and the Separateness of Propositions" involve violations of this restriction. Unsurprisingly, given the title: "Epistemic Teleology *and the Separateness of Propositions*".

Yes: reasons for believing P cannot be factors that bear on the truth or falsity of some logically or probabilistically unrelated proposition Q. That is a consequence of a restriction that flows from a general account of what it is for a consideration to be a right-kind reason for a commitment-involving attitude. Does this restriction show every kind of Strong Epistemic Teleology to be untenable? No. It does not undermine the version of Strong Epistemic Teleology with which I started in this chapter, which focused only on comparisons between the (probable) inherent properties of the doxastic alternatives with respect to one and the same proposition. That is why I consider it. Other versions fail on general grounds. We can see in advance that these theories collapse the distinction between right-kind and wrong-kind reasons.

As it happens, this is not far from what current theorists naturally labeled as epistemic teleologists are now doing. As I noted in §2.7 of Chapter 1, Goldman (2011) defends a reliabilist version of evidentialism. He is in effect willing to accept that justification for believing P turns on whether believing P would fit the total evidence one has that bears on whether P. He simply gives a process reliabilist account of what it takes for a belief to fit the total possessed evidence. As I argued before, this is as it should be. Evidentialism is best formulated as a meta-normative theory—a theory about relations *between* normative facts. Evidence is clearly normative. Only a conflation of providers of evidence with evidence itself could convince one otherwise. Evidentialism could be understood as a structural view that spells out what right-kind reasons for belief can be: only evidence bearing on whether P can constitute a right-kind reason for believing P.[184] Some of the most plausible theories of the right-kind/wrong-kind distinction entail this conclusion.[185] Epistemic teleologists like the current Goldman should be seen as taking the stage *after* high-level structural constraints of this type are imposed. Their views are about what it takes for right-kind reasons for belief to be sufficiently good, *granting* that they are always evidence. Happily, sophisticated theorists now in effect make this claim.

---

[184]As defended by Nishi Shah, this in effect *is* how evidentialism is understood. Of course, as I noted in Chapter 1, Conee and Feldman may not accept this understanding. So much the worse for them.

[185]This is clear with Parfit (2011) and Piller (2006), who analyze the right-kind/wrong-kind distinction, at least for commitment-involving attitudes, as the object-given/state-given distinction. But it also goes for Hieronymi (2005)'s theory. Indeed, anyone who thinks an essential mark of a right-kind reason is that one can *directly form the attitude for which it is a reason* simply upon recognizing its presence (and the absence of competing/undercutting reasons) will have grounds for denying that anything other than evidence and *perhaps* some pragmatic factors can be right-kind reasons for doxastic attitudes. Cf. Schroeder Ms_1 and Ms_2.

This is part of why Berker's putative counterexamples leave untouched the program he is looking to undermine. The program can be restricted in a perfectly principled way that honors the separateness of propositions. The restriction is dialectically reminiscent of a restriction that any reasonable consequentialist will want when she turns from giving a theory of reasons for *action* to giving a theory of reasons for *intention*. As formulated at the outset in this chapter, Strong Epistemic Teleology *honors the separateness of propositions*. It is an interesting teleological view all the same. Indeed, it is closer to the view that people in the literature implicitly accept than any version Berker considers. The implicit view that unites these people still needs to be considered, given its weighty implications for the deontic significance of epistemic rationality.

It is worth acknowledging that in even more recent work in progress, Berker points to a different kind of putative counterexample that does not turn critically on the separateness of propositions.[186] In his main new case, Jane Doe has a terminal illness but stubbornly believes that she will live significantly longer than the doctors predict. As it happens, her possessing this unjustifiedly optimistic belief greatly increases the chance that she will live longer. The belief makes a self-fulfilling prophecy. This case may seem more pressing. Here, after all, there is *proposition-relative* advancement of epistemic goals with clear absence of epistemic justification.

Nevertheless, the general kind of strategy I've been recommending—the strategy of appealing to meta-normative constraints that are antecedent to first-order theorizing and so available to Strong Epistemic Teleologists—can be invoked again. There are two options worth noting.

Firstly, notice that because we are examining a belief Jane Doe actually holds, our intuitions are primarily tracking *doxastic justification* if they are tracking any kind of epistemic justification. As I argued in Chapter 1, being doxastically justified requires not just that one's motivating reason for believing be a sufficient epistemic reason, but also that one believes for that reason in what I called the strong sense. This requires that, in using that reason as one's reason, one *tracks* what makes it a good epistemic reason. "Tracks" here means that using that reason as one's reason *manifests a (localized) competence* one has to respond to good epistemic reasons bearing on whether to believe in circumstances like the one in question, and *only* to such considerations in theoretical deliberation on such questions. Now, to *test* whether someone fails to believe for a sufficient reason in this strong sense, we can check whether the following counterfactual holds:

> if the relevant reason failed to be a good reason, the person would still believe for that reason.

This counterfactual holds in Jane Doe's case, at least as I understand it. If her belief weren't self-fulfilling and it indeed made it *less likely* that she would live longer, Jane Doe would still continue to believe, with no change in her doxastic agenda. So Jane fails to be doxastically justified. Notice that one cannot reply by trying to fix up the case. If Jane holds this belief because and *only* because it *is* self-fulfilling, it is far less plausible that she believes unjustifiedly.

Notice that this is something a Strong Epistemic Teleologist is permitted to say. A Strong Epistemic Teleologist can accept the view that being doxastically justified in believing P just is believing for sufficient epistemic reasons in the relevant strong sense. This, as I argued in Chapter 1, is a second-order, meta-normative theory of epistemic justification that leaves open what the criteria of quality for epistemic reasons are. The Strong Epistemic Teleologist is simply offering a first-order account of what makes epistemic reasons good or sufficient. This theory is

---

[186]His newer working paper is "The Rejection of Epistemic Consequentialism." Many thanks to Selim for sending me a copy and also for helpful correspondence on his earlier paper. While I disagree about points of strategy and the structure of the dialectical space, I am obviously in favor of his overall ambitions.

compatible with the view that doxastic justification just is believing for a sufficient reason in the relevant strong sense. Indeed, it would be unfortunate—and surprising—if it weren't.

There is a second move the Strong Epistemic Teleologist can make, this time at the level of propositional rather than doxastic justification. Another plausible structural constraint that a reasonable theory of epistemic justification should accommodate is that only *possessed* epistemic reasons can contribute to propositional justification. What the Strong Epistemic Teleologist should say is this. Jane simply does not possess a good reason for her belief. The good reason that Strong Epistemic Teleology correctly predicts to exist in this case is:

> the fact that forming the belief will increase the objective probability that the belief is true.

While the Strong Epistemic Teleologist should say that there *exists* a powerful reason for her belief—namely, this fact—Jane does not *have* this reason. Indeed, the most natural way of understanding Berker's case is that Jane has no epistemic reason for the belief in question at all. She believes just because it makes her feel good. There is no reason why an epistemic teleologist cannot agree that these facts prevent her from being justified, even propositionally. This is for the same reason noted earlier. She can accept the view that being propositionally justified requires having sufficient epistemic reasons. Her view comes in after this meta-normative theory has been set up. Her view provides a criterion of quality for epistemic reasons, which bears on the character of epistemic justification through that antecedently held meta-normative theory.

Notice, by the way, that act consequentialists are entitled to say the same kind of thing. Suppose the 9-digit combination to some safe containing a billion dollars is Dave's telephone number. Dave can either try to guess the number and be slapped a billion times if he doesn't get it within ten minutes, or be given 10% of the contents if he does 500 jumping jacks instead. He has no reason to think the combo is his phone number, nor other clues. Now, the fact that N = Dave's cell phone number *is* a powerful reason to dial N on the safe. Plausibly, Dave doesn't *have* a powerful reason to try to dial any number. He should do the jumping jacks *even if*, as it happens, *the number he is objectively most likely to guess first is his phone number*—say, because that is the number he is always disposed to guess first when asked to guess a 9-digit number. (He is highly egotistical, let's imagine.) That conclusion can be derived from plausible versions of the Factoring Account of having reasons—say, a version on which having R as the reason it is requires being in a position to know that R is the reason it is. There is no reason why an act consequentialist can't agree about this conclusion. Accepting the Factoring Account is not a retreat to something less than act consequentialism. The view at hand is an attempt to *embed* act consequentialism within a plausible meta-normative account of what it takes to have a reason.[187]

So I remain unconvinced by Berker's attack on Strong Epistemic Teleology. All the same, I think we should oppose sophisticated versions of epistemic teleology like the versions I formulated at the outset. These theories are powerful, interesting theories.

2.   Undercutting the Motivations for Weak Epistemic Teleology

I now turn to prepare the grounds for my attack. The first thing I want to do is ask why it may feel obvious that there *are* ultimate epistemic goals or intrinsic epistemic values that we ought to promote, for their own sakes or otherwise. I will not deny that there is a sense in which this is

---

[187]To avoid circularity, the Strong Epistemic Teleologist will have to be careful about how she understands the possession of reasons. But this, as I argued in Chapter 1, is a problem for everyone. Everyone should think that justification turns on possessed reasons. Everyone will then face the challenge of analyzing possession in a way that does not result in circularity. Return to §2.9 of Chapter 1 to see some options.

true. What I will deny is that it is the relevant sense needed for supporting Weak Epistemic Teleology. Seeing why this is true will prepare a more direct objection to Weak Epistemic Teleology. I'll turn to it at the end of this section, and develop it in full detail in the next section.

2.1.   *Goals and values that are epistemic* versus *epistemic goals and values*

I start with a small point. It will naturally bring us in the next subsection to a bigger one. Part of the appeal of thinking that there are ultimate epistemic goals or teleological epistemic values in some sense that could *actually support* Weak (and Strong) Epistemic Teleology rests, I believe, on a conflation of two types of claim. The following pair of claims illustrates one type:

> (Predicative-G)   There are ultimate goals of ours that are epistemic.
> (Predicative-V)   There are intrinsically valuable states that are epistemic.

Another type is illustrated by this pair of claims:

> (Attributive-G)   There are ultimate epistemic goals of ours.
> (Attributive-V)   There are intrinsically epistemically valuable states.

These are importantly different types of claim.[188] To illustrate this, I'll borrow an analogy from Ridge (forthcoming), who stressed the same difference for other reasons. Consider saying that

> (1)   the attack was strategically good,

and saying that

> (2)   the attack was good, and moreover strategic.

(1) can be true even while (2) is false. Morally abhorrent attacks can be strategically ingenious. A converse pattern holds too. Consider the claim that

---

[188]Michael Ridge has stressed the importance of this distinction for recent debates about the value of knowledge and understanding in his (forthcoming). While the idea occurred to me independently in running a discussion group on epistemic value, I acknowledge his influence now that I have read him. I would also like to note (with Ridge) that Ernest Sosa has been more careful about the distinction than others, though he has not discussed it under this description. See Sosa (2007: Ch.4) and Sosa (2010: Ch.3).

A month or so after reading Ridge, I discovered a sadly forgotten paper by Casey Swank that also makes something close to this distinction. Here is Swank (2000: 196–7):

> We might first move, with others before us, from the trivial observation that

>> ev1   There is something specifically epistemic about an epistemic vice.

> to the not so trivial conclusion that

>> ev2   An epistemic vice is a trait that is bad in a specifically epistemic way

> —bad, as ev2's proponents would say, from the point of view of truth, or of attaining truth while avoiding error, or something like that. But that would be a mistake: ev2 does not follow from ev1. If it did follow, we should have likewise to conclude that

>> ts2   A tall stranger is a person who is unfamiliar in a specifically tall way.

> For it is of course true that

>> ts1   There is something specifically tall about a tall stranger.

> But ts2 is false: There is nothing specifically tall about the way in which a tall stranger is unfamiliar.

(3)   the attack was strategically bad,

and the claim that

(4)   the attack was bad, and moreover strategic.

(4) can be false while (3) is true. With enough pacifists conscripted against their wills, the military may have deliberately ineffective strategies. Still, assuming the pacifists are right, this may be a good thing. Perhaps they damage nothing but abandoned buildings.

This reveals a crucial point. Suppose we are trying to write a book of military strategy. The book could be correct even if wholesale pacifism were correct. Conversely, even if wholesale pacificism were incorrect, that would be irrelevant to whether our book was correct. An analogous point holds for epistemology. It does not matter whether there are goals or teleological values that happen to be epistemic, if what we're interested in is whether anything is good in a teleological way from the epistemic point of view in a sense that is parallel to "good from the point of view of strategy". Moreover, it does not matter whether there are *ultimate* goals that happen to be epistemic, if what we are interested in are properly epistemic goals.

To be sure, what is valuable *simpliciter* and epistemic may place *indirect* constraints on epistemological theory. We do not want a theory of knowledge that, together with uncontroversial axiological premises, yields the conclusion that knowledge can be no more valuable than true belief. Still, an account of what is valuable *simpliciter* and epistemic will not *as such* constrain our view of what is *epistemically good* or of what we *ought epistemically to believe*, in the senses that belong properly to epistemology rather than to general axiology or deontology.

This matters. And many may claim to agree. After all, many acknowledge that Tom Kelly has decisively undermined the view that epistemic rationality is a restricted form of instrumental practical rationality, and the failure of that unfortunate view turns precisely on the distinction I am stressing.[189] Nevertheless, closely related mistakes persist. Consider Alston (2005: 30–1):

> I don't know how to prove that the *acquisition, retention, and use of true beliefs about matters that are of interest and/or importance* is the most basic and most central goal of cognition. […] But I suggest that anyone can see its obviousness by reflecting on what would happen to human life if we were either without beliefs at all or if our beliefs were mostly false. […] [T]he idea that it is important for human flourishing to be guided by correct rather than incorrect suppositions about how things are, where this is of interest or importance to us, is so obvious that it would seem to be unnecessary to belabor the point. […] I will say only […] that where we seek to produce or influence one outcome rather than another, we are much more likely to succeed if we are guided by true rather than false beliefs about the likely consequences of one or another course of action. That is the basic practical importance of truth. And again, with many other philosophers, I take it that there are more purely theoretical reasons for positively evaluating truth. […] [T]he attainment of knowledge and understanding are also of intrinsic value. "All men by nature desire to know," said Aristotle, and this dictum has been reaffirmed by many of his successors. Members of our species seem to have a built-in drive to get the truth about things that pique their curiosity and to understand how and why things are as they are and happen as they do. So it is as close to truistic as we can get in philosophy to take truth as a good-making characteristic, and falsity as a bad-making characteristic, of beliefs […].

---

[189]See Kelly (2002, 2003). While Kelly has encountered some opposition (see e.g. Leite (2007)), it is fair to say that the opposition fails to understand the key distinctions on which Kelly's points rest, which are partly imported from the work of people like Parfit and Piller in the philosophy of practical reason. This is brought out in Kelly's replies (see e.g. Kelly (2007)).

Alston rightly claims that true belief is of practical importance and also valuable as an end itself, simply because we are curious beings by nature. While this is true, it does nothing to motivate his teleological evaluation procedure for epistemology. An evaluation procedure that properly belongs to epistemology rather than general axiology will be one that explains what is *epistemically good*. It will be partly teleological if there are constitutively epistemic ultimate goals which are needed to explain some of the facts of epistemic value. Noting that true belief, even of interest, is a sensible goal of ours—indeed a sensible goal of ours *for its own sake*—is irrelevant. So the transition from the first sentence in the quote from Alston to the second is a *non sequitur*. What determines whether there is a goal of cognition is going to be some essential fact about cognition as such. Yet there could be cognizers who are not curious in the way we are—say, who address questions in an entirely perfunctory way—and even for whom true belief was of little practical importance, owing to an easy environment, created by some friendlier sibling of Descartes's demon. These beings could have attitudes that are evaluable from the epistemic point of view in the same way ours are, even if they shared neither our curiosity nor our needs.

Given the distinction I'm pressing, this should not come as a surprise. Pacifists might be right that we ought not to value military goals. Ambitious pacifists—say, the Kant of "Perceptual Peace"—may have even thought that our rational obligation not to have these goals flows from deeper aims that we have essentially as beings of the kind we are. Even so, there would still be such a thing as what is ultimately good from the point of view of military strategy. There would still be appraisals of strategies as militarily good or bad. Pacifists could agree. The same point goes for what is epistemically good/bad, when compared with what is good/bad and epistemic.

While this may seem obvious, Alston's conflation is pervasive. To take one illustration, recall how it is now standard in discussions of the "truth goal" for people to restrict it to questions that are *of interest* or that *pique curiosity* or that *are significant*. This in the Alston passage, and one finds it in Goldman (1999, 2002), Lynch (2004), Whitcomb (ms), and others. Insofar as the truth goal is supposed to be an *epistemic goal* and not just a *goal simpliciter that happens to be epistemic*, it is unclear what motivates the restriction. Indeed, as Grimm (2009) rightly stresses, the restriction makes it *harder* for epistemic teleologists to explain the distinctive normative force of facts of epistemic justifiedness and unjustifiedness. Even one who is interested in "trivial" subject matter, like the contents of a telephone directory, can be more or less justified in her beliefs about it, and to the same degree that you or I can be justified in something more interesting and important. Restricting the scope of the truth goal to questions of importance or of interest renders that putative goal irrelevant to explaining why these trivial beliefs are or are not justified.

Indeed, there is a further very obvious objection to this approach that has gone unnoticed to my knowledge. Consider the negative epistemic goal of not having false beliefs. This goal is more plausible as a properly epistemic goal, though I will later deny that it could be fundamental. Should this goal be restricted to questions of importance or interest, so that there is nothing epistemically bad about having false beliefs as long as these beliefs are about uninteresting subject matters? Surely not. Moreover, many of our beliefs simply aren't about interesting subject matters. Consider many of our basic perceptual beliefs, which we routinely form simply upon opening our eyes and looking around. I find it hard to see how these beliefs are necessarily significantly more interesting than beliefs about blades of grass or random telephone numbers. So restricting the goals to apply only to interesting subject matters would simply put these beliefs outside of the scope of evaluation. That is clearly a mistake.

This style of approach ignores the distinction I've been pressing. The epistemic normative force of facts of justifiedness and unjustifiedness is insensitive to how interesting or important the subject matter is, except perhaps where *stakes* are concerned, which may affect the *sufficiency* of reasons for belief. Moreover, the epistemic badness of an unreliable belief-forming process is not

lessened if that process is only capable of outputting uninteresting beliefs, which many of our basic beliefs—e.g., our perceptual beliefs—surely are. It is far better to drop the restriction to topics of curiosity or importance, which rests on a conflation. It is better to focus explicitly on properly epistemic goals rather than goals that happen to be epistemic—even final goals that we have by being curious beings. Only a conflation of these two senses of "epistemic goal" could encourage this mistake. Once the mistake is pointed out, it is unclear whether there even are properly epistemic goals, at least of fundamental importance for epistemic evaluation.

Of course, someone on the other side might try to claim that it will be *easier* to show that true belief is a constitutively epistemic ultimate goal—an ultimate epistemic goal in the proprietary sense associated with claims like Attributive-G. Consider the claim that true belief is one of our ultimate goals *qua* epistemic subjects. Don't many people believe this? Isn't it at the heart of discussions of the "aim of belief", understood as a *constitutive aim*? If so, there may seem to be a simple way in which true belief is a properly epistemic, ultimate goal of ours.

But this reasoning is confused for two reasons. To bring out the first, it is worth stressing again that we have no epistemic obligation to form beliefs. We are not defective *qua* epistemic subjects for lacking doxastic attitudes about virtually anything, simply because we never consider questions, or don't want to form doxastic attitudes rather than less commitment-involving cognitive attitudes like the weaker attitude of *acceptance* discussed by Bratman, L. J. Cohen and Wright. The impulse to form beliefs in the first place is a practical one, in a broad sense that includes goals that *happen to be epistemic*, such as quenching curiosity. Once we have a belief, we prefer that it be true and we prefer not to have it if it is false, *qua* epistemic subjects. *These* preferences are indeed constitutive ideals of epistemic agency. But none of this is sufficient to show that true belief is a constitutively cognitive ultimate goal of ours. *Believing* is a practical goal, in a broad sense that includes aims like quenching curiosity (aims not, as I noted before, shared by all cognizers as such, and not *epistemically culpably* so). *Given* the goal, our essential nature as cognitive beings forces us to want to persist in believing only if our belief is true. It does not follow that our desire for true belief is essential to our identities as epistemic agents.

There is a general fallacy here.[190] The only way to get the conclusion that true belief is an ultimate constitutively epistemic goal would be to engage in a piece of reasoning with this form:

> B-ing is a goal of ours.
> As an essential fact about us *qua* A-ers, we want to be B-ing only if we B C-ly.
> _____
> So, B-ing C-ly is a constitutive ultimate goal of ours *qua* A-ers.

This reasoning fails when B-ing *simpliciter* is not an ultimate constitutive goal for A-ers. As an illustration, suppose Billy is a committed member of some cult. Being so committed demands that if one wants to experience some *objet d'art*, it must be one that celebrates the cult. Suppose now that Billy wants to listen to music just because music is something he values for its own sake. So as a committed member of the cult, he wants to listen to cult music. It hardly follows that it is an essential fact about him *qua* committed cult member that he want to hear cult music for its own sake. If we have no epistemic obligation to form beliefs, true belief formation stands to us *qua* epistemic agents as cult music appreciation stood to Billy *qua* cult member.

For those of us who belong to the cult of inquiry, this may feel unsatisfying. Perhaps it would be nice to be able to condemn those who take no interest in the life of the mind simply on the basis of epistemic value theory. Yet epistemic evaluation is not applicable only to people who value

[190]Something close to this point is made in Sosa (2001) and (2003).

inquiry as such. We can evaluate the cognitive attitudes of people who don't. These people can be highly epistemically virtuous, in the relevant proprietary sense. One is no less careful a reasoner or conscientious a thinker or successful a believer *simply* by failing to love inquiry as some academics do. Perhaps there is some contingent correlation. But surely a weak one. One has a defect, perhaps, in the sense that one misses out on something of intrinsic value. But not a defect of the kind that first-order normative theories in epistemology should ban. True first-order normative theories in epistemology cannot enjoin us to be intellectuals.

True belief, then, is not a constitutive ultimate goal of ours simply *qua* cognitive subjects. At most, only the final *conditional* goal of believing *only if* true fundamentally characterizes us *qua* cognitive subjects. Even the final conditional goal claim may be mistaken as a description of what fundamentally characterizes us *qua* cognitive subjects.

I'll argue that it is later. For now, I will just note how this *could* be mistaken. The key point is this: fundamentally, truth may just be a standard of belief *to be respected* by all cognizers as such. Its being such a standard could explain why the conditional goal is a *derivative* one at which we *must aim*. After all, it is hard to see how one could have sufficient respect for the truth without having the aim of ensuring that you believe P only if P. Crucially, however, it would not support the stronger claim that the conditional goal is *the* (or *a*) *fundamental* goal. What is fundamentally a *standard* is not (clearly) fundamentally a *goal*. Fundamentally, it is rather a constraint to be respected. But if *this* is true, Weak and Strong Epistemic Teleology both fail. If truth is in the first instance just a standard to be *respected*, sometimes we ought epistemically *not* to promote the goal. Why? Because respect for the truth consists partly in satisfying requirements of rationality. Rationality can require otherwise, as we know from Chapter 3. If *respect* for the standard is more fundamental, it can trump promoting the goal. I'll argue for this in §3.5.

The intuitive data so far underdetermine which of these explanations of why the conditional goal is <u>a</u> goal of ours *qua* epistemic agents is correct. It is of great importance whether it is a bedrock fact about us that we have the goal of believing only if true *qua* epistemic subjects—so that this is an *ultimate* goal—or whether we aim at this goal *qua* epistemic agents only because we ought epistemically to respect truth as the standard of belief formation. If the latter is true, and the conditional goal is aimed at only derivatively, requirements of epistemic rationality may, as I'll argue in Chapter 6, have robust deontic significance. And as I'll go on to argue later in this chapter, further intuitive data plus principles of theoretical parsimony favor my alternative order.

In any case, what matters right now is this: true belief itself is not an ultimate goal of ours *qua* epistemic agents. We do not, simply *qua* epistemic agents, seek more instantiations of it, even interesting instantiations of it. If we do, we do only for practical reasons. Again, I understand "practical reasons" broadly. Our interest in some questions may be intrinsic and so worth calling "purely epistemic" in a *predicative* way. But that is still essentially practical: intrinsic curiosity is a *desire* that we have. We may value some true belief for its own sake, so that it shows up on a Moorean list of values, along with pleasure and beauty. This may be true because we are *properly* curious beings. It would not follow that true belief is an ultimate epistemic goal in the *relevant* sense. For in the imagined scenario, true belief is on the list of *ultimate goods*—the list of things that are *good simpliciter*. Nothing follows of relevance to the status of epistemic teleology.

Moreover, it is easy to imagine people who do not value true belief or knowledge for its own sake, and who are not so curious. If it or knowledge is on the Moorean list, there is something *broadly* defective about them. But only in the same sense in which there would be something broadly defective about them if they failed to love beauty. They can still be epistemic subjects, subject to epistemic evaluations in the same way that lovers of truth like we philosophers are.

So I think there is a burden of proof on one who wants to regard true belief, or any unconditional end state, as a properly epistemic ultimate goal of ours *qua* cognizers. Without categorical epistemic pressure for forming commitment-involving cognitive states—pressures I don't see— the most that can be said is that we have the conditional goal of believing only if true *qua* epistemic subjects. But that is an importantly different claim. And as I've stressed, even this is not sufficient to support epistemic teleology in any form. Everything of importance turns on whether this goal is an *ultimate* goal of ours or whether we only must aim at it because of some more basic fact, such as the potential fact that *qua* epistemic subjects we must see ourselves as being under an obligation to respect truth as the standard of belief formation.

Again, if the latter hypothesis is true, neither Strong nor Weak Epistemic Teleology succeeds. Respect for the truth may require complying with requirements of epistemic rationality. As we in effect saw in the last chapter, requirements of epistemic rationality can require *systematically frustrating* even the conditional goal when the ultimate nondoxastic appearances are sufficiently misleading. So respect for the truth may in fact require "doing" what would frustrate that goal.

2.2.   *The intrinsic/extrinsic distinction* versus *the final/instrumental distinction*

There is a deeper reason to doubt the motivations for Weak Epistemic Teleology. It supports my alternative explanation of why we derivatively have the epistemic aim of believing P only if P.

Weak Epistemic Teleology implies that all epistemic value is either final or instrumental epistemic value. While many epistemologists may not recognize it, this is an extraordinarily controversial implication. Of course, it can sound uncontroversial if one simply presupposes that

> (Intrinsic=Final)   all intrinsic value (*simpliciter* or domain-relative) is *final* value, or value to be brought about for its own sake,[191]

and that

> (Extrinsic=Instrumental)   all extrinsic value (*simpliciter* or domain-relative) is instrumental value, to be brought about as a means to finally valuable results.

But these claims are widely recognized to be false in the broader axiological literature. There are many stock examples that illustrate why these claims are false.[192] It would be surprising if the

---

[191]Some may use "finally valuable" to mean "worth wanting for its own sake". Following Korsgaard (1983) and many others after her, I use "final" so that it coordinates with "instrumental". The finally valuable is that to which the instrumentally valuable is conducive (causally or constitutively). The terminology itself doesn't matter for the point I'm going to make. (The jargon is partly philosophy's invention in any case.)

   I should note that when applied domain-relatively, "to be brought about" in (Intrinsic=Final) and (Extrinsic=Instrumental) is also to be read in a domain-relative way. Infinitives can of course be read normatively ("The good is to be promoted" means "The good ought to be promoted"). While I am not apprised of the current semantics, they can doubtless be implicitly restricted to a particular domain.

[192]See for instance Korsgaard (1983), Anderson (1993), Rabinowicz and Rønnow-Rasmussen (1999), Dancy (2004a), Langton (2007) and Tannenbaum (2010), to list a few of the most salient works in which such examples are discussed. Ironically, the very same Shelly Kagan, whom Scanlon cites in my epigraph as equating the good with what is to be promoted, helped this in fact deeply opposing cause in his (1998). Perhaps he simply changed his mind over the years. Scanlon was citing Kagan (1989: 60).

   Some people in the epistemic value literature are aware of some of this literature. Pritchard (2011), e.g., cites some of it. As I'll make vivid in the next chapter, these people have not, I believe, appreciated how important the underlying distinction is. It is the key to defending a veritist monism about intrinsic epistemic

reasons why these claims fail did not extend to properly epistemic value theory.

Rather than piling up examples, I am going to point to a theoretical argument against these claims in their most general forms. It will turn out to be key to seeing why Weak Epistemic Teleology fails.[193] I'll discuss the non-epistemic case first to illustrate how claims like this clearly fail, and to raise indirect doubt for the restricted application of the general claims to epistemic axiology. Along the way, I'll point to a crucial analogy for the epistemic case. It will provide a stepping stone for a direct argument against Weak Epistemic Teleology that I will then erect.

To bring out the theoretical argument, I'll start by noting that everyone needs to acknowledge that there is some fundamental connection between (a) being valuable and more determinate evaluative properties (even within some domain), and (b) correct *pro-responses* or, more helpfully, *ways of valuing*. The latter include both positive *acts* like promoting, preserving and protecting, and positive *attitudes* like seeking, loving, admiring, appreciating, feeling awe at, being amused by, being attracted to, intending to produce, and desiring to be so. This claim is in part made plausible by the sheer structure of words that are often used to identify determinate evaluative properties—*admirability*, *attractiveness*, *desirability, awesomeness*, *amusingness*, etc.—which wear pro-responses on their sleeves. But it is also made plausible by what is involved in *valuing* something, even relative to a domain— e.g., moral, aesthetic, epistemic. As Anderson nicely puts it: "To value something is to have a complex of positive attitudes toward it, governed by distinct standards for perception, emotion, deliberation, desire and conduct."[194] And as Scanlon said, to understand the value of something is to *know how to value it*. Things that have positive evaluative properties are necessarily *fitting or correct objects of valuing* (*in some way of valuing or other*).[195]

A more specific connected claim that matters for the purpose of seeing why Intrinsic=Final and Extrinsic=Instrumental are controversial is the following:

> (**Fundamental Link**) For anything valuable (perhaps domain-relatively) or that possesses some determinate evaluative property, there a *way* (or *some ways*) *of valuing* it—a pro-response R or set of pro-responses {R$_i$}—whose correctness is *fundamentally linked* with its being valuable (perhaps domain-relatively) or its having that determinate evaluative property.

"Fundamentally linked" is the important qualifier. To illustrate what I intend, consider the specific evaluative property of *desirability*. The pro-response that is *fundamentally linked* with desirability is *desiring*. There are certainly other pro-responses that are appropriate to desirable things. Sometimes we ought to ensure that desirable things continue to exist, and so *protect* them. Still, the pro-response of protection is not *fundamentally* linked with desirability. It is *derivatively* linked. Often, something is worth protecting partly because it is antecedently desirable.

A final qualification will bring us to the key point. While it is often clear what pro-responses are fundamentally linked with specific evaluative properties like desirability, it is not clear what the fundamentally linked pro-responses are in the highly generic case of something's simply *being*

---

value. Pritchard (2011)'s view that monism vs. pluralism about epistemic value is particularly key to disputes about the value of knowledge and understanding rests, I believe, on a failure to appreciate the importance of the distinction. I'll argue at length for this in a later chapter.

[193]The argument is not new, just the application to epistemic value. Cf. Anderson (1993) and Scanlon (1998).

[194]Anderson (1993: 2).

[195]It is worth stressing that the thought that there is such a connection is not special to *buck-passers* about the evaluative. Buck-passers claim that something's having an evaluative property *reduces to* its having further properties that make certain ways of valuing correct or fitting. Buck-stoppers, on the other hand, claim that the presence of evaluative properties *explains why* valuings of things would be correct or fitting. Even buck-stoppers (should) see an essential connection between values and fitting/correct pro-responses.

valuable (even domain-relatively). What remains clear is that there will be some *priority structure* among the pro-responses connected with something's simply being valuable. Here is an illustration of how priority structure could arise. Some valuable things are worth preserving. But preservation many not be the correct pro-response that is fundamentally characteristic of the value of these things. Why not? Because it is fairly intuitive that something is *worth preserving* only if was *worth creating in the first place.* This may turn out to be wrong, but I hope the reader will find some *prima facie* plausibility in this case, which I mention only for illustration. What matters for our purposes is that this priority structure will *bottom out* in particular cases where particular things are valuable, so that there are some pro-responses whose correctness explains the correctness of any other correct pro-responses to something of value. This is primarily what I will have in mind in talking about pro-responses that are *fundamentally connected* with simply being valuable (even domain-relatively) in what follows.

Now, in the sense that that coordinates with "instrumental", "final" value by definition is value *as a goal to be brought about for its own sake.* Instrumental value by definition is value *to be brought about* as a means to such a goal. Given the Fundamental Link, we can now see that to presuppose Intrinsic=Final and Extrinsic=Instrumental is to presuppose an extremely restrictive picture of the pro-responses that are fundamentally linked with value (i.e., that are at the bottom of the priority structure among correct potential pro-responses). The fundamentally correct pro-response to both final and instrumental value—value as a goal or end—is to *bring it about*, *as an end or means.* Yet perhaps not everything intrinsically valuable is such that the fundamentally correct response to it is to bring it about as an end in itself. And perhaps not everything extrinsically valuable is such that the fundamentally correct response is to bring it about as a means. Intrinsic=Final and Extrinsic=Instrumental thus make a highly controversial predictions about how the priority structure among correct pro-responses to value will bottom out.

One doesn't need to reflect on examples to appreciate how controversial this is. There are *many* ways of valuing whose correctness might be fundamentally characteristic of different values, intrinsic or extrinsic. Consider Nozick's list of ways of valuing: "bringing about, maintaining, saving from destruction, prizing, contemplating […] car[ing] about, accept[ing], support[ing], affirm[ing], encourag[ing], nurtur[ing], protect[ing], guard[ing], prais[ing], seek[ing], embrac[ing], serv[ing], be[ing] drawn toward, be[ing] attracted by, aspir[ing] toward, striv[ing] to realize, foster[ing], express[ing], nurtur[ing], delight[ing] in, respect[ing], be[ing] inspired by, tak[ing] joy in, resonat[ing] with, be[ing] loyal to, be[ing] dedicated to […] be[ing] elevated by, enthralled by, lov[ing], ador[ing], rever[ing], be[ing] exalted by, be[ing] awed before, find[ing] ecstasy in."[196]   The assumption that bringing about an end in itself is the response that is fundamentally characteristic of intrinsic value cries out for support. It is unobvious and frankly arbitrary.

Of course, a friend of Intrinsic=Final and Extrinsic=Instrumental can *try* to explain away contrary appearances. For a value that seems intuitively to call in the first instance for *respect,* she can try to say that what is *really* valuable is just the state of affairs of X's *being respected*, or the mental states involved in respecting X. But it is highly intuitive that this gets things completely backwards. Respect for X often deserves to be brought about *because X is antecedently valuable* in the particular way of being something that merits respect. That suggests that the state of affairs of respecting X—as well state of affairs of causing X to be respected—has *extrinsic* value. It has value that is *derived* from the prior value that X itself possesses, as a value to be respected.

That is a fundamental problem with collapsing the intrinsic/extrinsic distinction into the final/instrumental distinction. It predicts the wrong priority structure among pro-responses in many cases. It fails to appreciate the plurality of ways of valuing whose correctness could be

---

[196]Nozick (1981: 429–30).

fundamentally bound up with particular valuable things.

A further vivid illustration of the problem is that the conjunction of Intrinsic=Final and Extrinsic=Instrumental implies that only *events* or *states* can have intrinsic or extrinsic value, since only events and states are the primary relata of the *bringing about* relation. Yet things and people and properties can obviously be valuable. A teleologist may try to say that it is only the *existence* of things and people that is basically valuable, or only the *instantiation* of properties that is basically valuable. We can, after all, bring about the existence of things and people, and cause properties to be instantiated. But further reflection shows this to be a deeply implausible maneuver. I find what Parfit (2011 v.1: 237) says convincing:

> According to some writers, [the view] can be widened to cover the goodness of some persisting things, such as people and works of art. Such things are claimed to be good when their nature gives us reasons to want them to exist, or continue to exist, and reasons to make that happen if we can. Moore even writes:
>
>> when we assert that something is good, what we mean is that its existence or reality is good.
>
> But these claims are mistakes. Something's existence can be good though this thing itself is not good, and *vice versa*. There are many bad people, for example, whose continued existence would be good as an end. When some good person is dying a slow and painful death, the continued existence of this person may be bad as an end. And there would be nothing good in the continued existence of good works of art if no one could ever see them.
>> According to what Scanlon calls *teleological* theories, it is only acts and other events that have *intrinsic* value in the sense of being in themselves good. Scanlon rightly rejects this claim. There are other things that can be in themselves good, such as people, books, and arguments. Since these things are not events, we cannot want them to happen, or make them happen. But we can respond to them in other ways. We can have reasons to read good books, be convinced by good arguments, and try to become more like good people.

It is far more plausible to reverse the Moorean claim that Parfit quotes, and to hold that states of affairs and events are often only valuable because of the prior value of their constituents. Anderson (1993: 26) makes this reversal plausible, partly by appealing to the connection between values and fitting pro-responses with which I started. Her points are worth quoting:

> [S]tates of affairs are generally only extrinsically valuable, because our intrinsic evaluative attitudes do not generally take them as their immediate objects. It makes sense for a person to value most states of affairs only because it makes sense for him to value people, animals, and other things. […] [S]tates of affairs that consist in the existence of something are valuable only if it makes sense to care about the thing that exists. It doesn't make sense to care about the existence of a painting unless it makes sense to care about the painting itself, perhaps because it is beautiful. And beauty is a valuable attribute of the painting, not of the fact that the painting exists. […] It may make sense for me to love a person, but this does not imply that I must want that person to continue living. If he is gravely ill, it may be the best expression of my love for him to wish that he die quickly and mercifully. A remarried widow may still love her long-dead husband, but be appalled if he were to pop back into existence.

As Anderson notes, if we single out values by looking at the objects of intuitively reason-supported "intrinsic" pro-attitudes—i.e., pro-attitudes held toward objects for their own sakes—we will notice time and time again that these objects are, well, *objects*, and not often states of affairs or events. Assuming that intrinsic values are necessarily objects of reason-supported potential intrinsic pro-attitudes, we will find a wealth of counterexamples to Intrinsic=Final.

These are a few reasons why Intrinsic=Final and Extrinsic=Instrumental are controversial. Given

that Weak Epistemic Teleology presupposes domain-restricted versions of these claims, we should find Weak Epistemic Teleology considerably less obvious.

Of course, so far I've only discussed examples outside of the sphere of properly epistemic value. There will be a positive burden on me to explain what the analogues are. I will discharge this burden. For now, the key point is that there is also a burden on the other side. The Weak Epistemic Teleologist owes us reasons for thinking that structural assumptions about the nature of intrinsic and extrinsic value that are false or at least very controversial in the practical sphere should be immune from criticism when we turn to the epistemic sphere.

2.3.   *A key illustration of how explanatory priority can reverse: the value of persons*

To bring out an analogy that will help with my explanatory burden, it is worth considering further one of the examples that Parfit and Anderson both identify—viz., the value of persons.

Persons clearly have intrinsic value, in the sense that they have value that *isn't derived* from anything more basic. Yet, as Jan Narveson once said with great force, while we are in favor of *making people happy*, we are neutral about *making happy people*. Understanding these observations is not easy on a view where the only response fundamentally called for by intrinsic value is *to bring it about*, as an ultimate goal or a means to one. Given Narveson's point, the intrinsic value of persons cannot consist in the fact that we ought *to make more of them* or even see to it that they exist for an indefinitely long period of time. Indeed, correctly responding to the intrinsic value of persons may call for different responses. Scanlon (1998: 104) masterfully illustrates this:

> Appreciating the value of human life is primarily a matter of seeing human lives as something to be respected, where this involves seeing reasons not to destroy them, reasons to protect them, and reasons to want them to go well. Many of the most powerful of these reasons, however, are matters of respect or concern *for the person whose life it is* rather than of respect *for human life*, or *for this instance of human life*, in a more abstract sense. The difference between these two forms of respect comes to the fore in cases of euthanasia and suicide.
>
> Suppose a person is in an irreversible coma. Would it show a lack of the respect called for by the value of human life to end this life by withholding food and other life supports, or to fail to protect it by providing protection against disease? Would a person who faces a life of endless unremitting and incapacitating pain show a lack of respect for his or her own life by seeking to end it? These questions are controversial, but I believe the answer in both cases is "No." This suggests that while appreciating the value of human life involves seeing that there are strong reasons for protecting life and for not destroying it, these reasons are restricted by the qualification "as long as the person whose life it is has reason to go on living or wants to live".

As Scanlon brings out, the failure of the application of teleology becomes vivid when we recognize that persons essentially call for respect, and that *this* is the pro-response more fundamentally linked with the distinctive kind of value they have. Can a teleological framework that collapses intrinsic and final value explain away the intuitions? Shall we say that what's *really* valuable in the first instance are the *acts of respect*, and that persons are only valued as means to these acts? Certainly not. The reverse is true: acts of respect have value only because persons have worth. Shall we say that what is really valuable is just the history of a person's conscious life? This is more implausible. Even persons whose conscious lives are terrible have worth. Such people are not *worth duplicating*. But they are loci of intrinsic value all the same. The only clear response available to the teleologist is a stubborn one. She must suppose that aim is to produce more instances of the conditional state of affairs of *people being well off if they exist*. This is more fundamental, captures the value of persons, and is not itself explained by anything more basic. Theorists like Scanlon will rightly say that this looks both arbitrary and implausible.

Whether or not people in ethics like Scanlon and Anderson are right about all this, the structure of their reasoning does uncover an important analogy. Notice that we want to replace:

(The Unconditional Goal)   We ought to cause more people to exist who are treated well, and who are well off in other ways.

with

(The Conditional Goal)   We ought to see to it that, for any person who exists, this person is treated well, and is well off.

More importantly, our desire to switch from the Unconditional to the Conditional Goal for persons is not just a *brute asymmetry* that axiology can't explain. It can and should be explained by a *reversal* in how we are to understand the priority structure among fitting or correct pro-responses connected with the value of persons. It is the fact that persons *more fundamentally* call for respect that explains the asymmetry between our reasons to cause more people to exist who are treated well and are well off, and our reasons to treat well and make well off people who already exist. This is unsurprising. If the fundamentally correct response associated with the value of persons *were* the teleological one of causing to exist—even in certain states, with certain properties—we would not expect this asymmetry. Since this asymmetry exists, we should expect a corresponding asymmetry in the priority structure among the correct pro-responses that are connected with the value of persons.[197] Otherwise we are left with an asymmetry that calls out for explanation, and that, moreover, can be explained without great difficulty.

The reasoning rests on a more general principle, which inherits plausibility from its usefulness:

> **Principle of Priority**: If there is a deep asymmetry in plausibility between an unconditional and a conditional goal claim connected with a value V, we ought to try to explain it by supposing that the correct pro-response that is *more fundamentally linked* with V is not the teleological one of *bringing about*, but rather some non-teleological one (e.g., respect).

We ought, that is, to *reverse the priority structure* among fitting or correct pro-responses linked with V, so that non-teleological responses are explanatorily prior to the restricted teleological ones.

Given the Principle of Priority, we can put together our observations from the last two subsections to provide a direct rebuttal of Weak Epistemic Teleology that is symmetrical to the rebuttal in the practical sphere, which Scanlon and others suggest. I will now turn to this.

3.   Rebutting Weak Epistemic Teleology

3.1.   *A direct argument from the conditional/unconditional goal asymmetry*

We have already seen that there is an asymmetry in how the "truth goal" is to be understood. Once we distinguish between goals that happen to be epistemic and properly epistemic goals, the following has no plausibility:

(The Unconditional T-Goal)   We ought epistemically to produce more true beliefs and fewer false beliefs.

---

[197]Whether this move explains everything is a question I cannot address here. I am only discussing the move for the sake of analogy. A complete evaluation of it would require addressing Parfit's Non-Identity Problem. Parfit's problem reveals contrary intuitions which can be used to motivate the introduction of values that are not "person-affecting"—values, that is, that are not focused on making things go better or worse *for anyone who will ever exist*. The ethical case is bafflingly hard to understand thanks to this problem.

This claim fails mainly because we have no epistemic obligation to form beliefs, and *a fortiori* no epistemic obligation to form true beliefs. As I stressed above, it is irrelevant to epistemic evaluation proper to note that we can restrict the Unconditional T-Goal by sticking "about subject matters of interest or importance" after "beliefs". Such qualifications are important if we are trying to understand values or goals *simpliciter* that happen to be epistemic. But we should avoid them if we are trying to understand epistemic value or goals proper. This becomes obvious when we think about false beliefs. Many of our basic beliefs—e.g., our perceptual beliefs about our immediate surroundings—are scarcely more interesting than beliefs about blades of grass or random telephone numbers. If these turn out to be reliably false owing to defects in one's basic belief-forming faculties, they are hardly *not as epistemically bad* or obstructive to the epistemic goal of avoiding false beliefs *just by being boring*. This is one illustration of why such restrictions are irrelevant to epistemic evaluation proper. But the more important point can be seen without examples: once we distinguish epistemic values or goals proper from values or goals that are epistemic, we should realize that the intuitions in favor of the Unconditional T-Goal (with pragmatic qualifications) are unreliable, and rest on illicit blurring of this distinction.

The following is more plausible as an expression of a properly epistemic goal:

> (The Conditional T-Goal)　　　　We ought epistemically to see to it
> that we believe P *only if* P is true.

But there is an important question. Is the Conditional T-Goal *fundamental*, so that the pro-response whose rightness is fundamentally linked with the epistemic value of truth in belief is bringing about—just *conditional* rather than *unconditional* bringing about? Given the Principle of Priority, the answer is "probably not". If so, Weak Epistemic Teleology is probably false. Indeed, given that we undermined the motivations for that view in the last section, I think we *should* conclude that Weak Epistemic Teleology is false, and start looking for an alternative.

This answer is appealing. It is odd to think that true belief *is* something fundamentally "to be promoted" even though there is a radical asymmetry between the plausibility of positing an epistemic obligation to produce more true beliefs and of positing an epistemic obligation to see to it that beliefs that we already have are true. This asymmetry calls out for explanation just like Narveson's asymmetry did. It is structurally identical: *making more true beliefs* stands to *making more happy people* as *making more beliefs true* stands to *making more people happy*. One might try to explain the asymmetry by denying that truth is an intrinsic epistemic value at all. One might also try to take the asymmetry as a bedrock fact. The first move would be a mistake. The second should be avoided unless we lack a clear alternative. It is better, then, to deny that the epistemic value of truth (in belief) is *fundamentally* to be understood in teleological terms. So we ought to look for a non-teleological explanation. This is like what Scanlon and others suggest we do to explain Narveson's asymmetry. By viewing persons as fundamentally to be respected, an explanation of the asymmetry was available.

What might the explanation be in our case? My favored explanation is an epistemic Kantian one, similar in spirit to the respect-based explanation of Narveson's asymmetry. On this view, the epistemic value of truth is fundamentally tied to its being a standard to be respected in the particular case. More specifically, the epistemic value of truth is fundamentally linked to

> the fact that, in looking to form a doxastic attitude on the question whether P, we ought epistemically to give the rule *believe P only if P* full weight in deciding the course of our theoretical deliberation about P,

since

giving this rule full weight in our theoretical deliberation *just is what it is* to respect the truth about whether P, and we ought to respect the truth about whether P.

On this view, we can *derive* the conclusion that we ought epistemically to constrain our doxastic deliberation by the rule expressed in the statement of the Conditional T-Goal *from* a more fundamental premise about what correct pro-response—viz., respect—is fundamentally linked with the epistemic value of truth. Notice moreover that we *cannot* derive the conclusion that we ought epistemically to constrain our doxastic deliberation by the Unconditional T-Goal from a more fundamental premise about what correct pro-response is essentially connected with the epistemic value of truth. So the epistemic Kantian view predicts that there should be an asymmetry in intuitive plausibility between these two goals by making truth's epistemic value out to be fundamentally non-teleological. The asymmetry in plausibility is not just some brute fact.

This is a significant virtue of the epistemic Kantian approach. Given the existence of the asymmetry in plausibility between the Conditional T-Goal and the Unconditional T-Goal and Principle of Priority, the epistemic Kantian approach is to be preferred to a watered down version of Weak Epistemic Teleology that requires taking the asymmetry to be a brute fact, and requires seeing us as basically aiming at producing more instances of the conditional state of affairs type <*if B is held, B is true*>. Given that we can cast doubt on the motivations for Weak Epistemic Teleology as suggested earlier, we can and should reject this view without scruple.

Of course, if there were other alternatives besides epistemic Kantianism that could be used to underpin the reversal of explanatory order suggested by the Principle of Priority, we wouldn't have a complete argument for this view. Nevertheless, we do so far have a direct argument against Weak Epistemic Teleology. That view posits an arbitrary normative asymmetry where there is in fact an asymmetry that *can be explained*—just not in fundamentally teleological terms.

3.2.   *Support for the epistemic Kantian reversal: patterns of epistemic value derivation*

Why accept the specific epistemic Kantian reversal I have recommended? One reason is that it correctly predicts some patterns of epistemic value derivation that we'd expect if truth were an epistemic value primarily to be respected—i.e., one to which the *fundamentally* correct pro-response is respect, and which is at most derivatively to be promoted, conditionally or otherwise.

To illustrate the thinking, consider non-epistemic cases first. When V is a value primarily to be respected, certain patterns of motivation that constitute respect for V will *by themselves* be *pro tanto* valuable because they constitute respect for V, whatever else their consequences.[198] Moreover, certain patterns of motivation that constitute disrespect for V will be noninstrumentally *pro tanto* disvaluable because they constitute disrespect for V. By contrast, when V is a value "to be promoted", such patterns of motivation won't by themselves be *pro tanto* valuable or disvaluable, except insofar as they happen to promote or eliminate V.

---

[198]One broadly teleological thinker accepts this idea—namely, Thomas Hurka in his (2001). Hurka, however, claims that it is *intrinsically valuable* to have pro-attitudes to things of intrinsic value. This claim rests on an unhelpful way of drawing the intrinsic/extrinsic distinction for value—admittedly, an unhelpful way that attracts others, like Michael Zimmerman in his (2001). Suppose I value friendship. My valuing it is good. But it is good *because* friendship is itself good. So my valuing friendship *derives its goodness* from the goodness of friendship. What is derivatively good in this way is worth calling "extrinsically good". So it is also worth denying that love of virtue is intrinsically valuable. It may be valuable "*considered in isolation*", or in virtue of its own *intrinsic properties* as a relational state with a certain object. But those are different, less useful ideas. If we're interested in the *derivation* of value and in what value is *underived*, we ought not to focus on them.

Take something that clearly has value primarily "to be promoted"—viz., one's own pleasure. Caring a lot about one's own pleasure and doing things that manifest such care *themselves* have no value—not even *pro tanto*—apart from their consequences.[199] (And the consequences are not likely to be good, given the infamous paradox of hedonism.) Similarly, not caring intensely or at all about one's own pleasure isn't itself disvaluable. It would be extremely unfortunate if it were! For then, given the paradox of hedonism, we could only achieve our pleasure by *failing to fully respond* to its value! The implausibility of this conclusion, among other things, should lead us to conclude that *only* the response of *bringing about more pleasure* is valuable. That is precisely because this is the pro-response *fundamentally connected* with the value of pleasure.

By contrast, caring about persons in a way that constitutes respect for them does in itself have *pro tanto* value, and failing to care about persons in a way that constitutes disrespect for them does itself have *pro tanto* disvalue. These are *pro tanto* virtues (/vices). Their status as *pro tanto* virtues (/vices) does not depend on consequences—not when even as viewed as *types* of attitude. Of course, if we learned that the best way to actually help people was *not* to care about them and disrespect them instead, we may have sufficient reasons *to cause ourselves* to disrespect them. Still, we would have reasons to care for them, and not to disrespect them. There is simply a distinction to be drawn here, as I noted earlier in §1.2. Caring about them would have the same *pro tanto* value by itself. Disrespecting them would have the same *pro tanto* disvalue by itself. This may be because the intrinsic value of persons is fundamentally to be respected.

The pattern we see here with the intrinsic value of persons is the same pattern we see in the epistemic case with the intrinsic epistemic value of truth, considered as a property of doxastic representations. Set aside *deontic* intuitions about demon worlds—set aside, in other words, the (questionable) intuitions that drive Cohen's new evil demon problem for reliabilism—and instead consider some simple *evaluative* intuitions. Suppose we live in a world with a wryly benevolent demon. This demon will reward reckless failure to care about the truth of one's beliefs. One will end up with mostly true beliefs and no false beliefs if one doesn't care a whit about the truth of one's beliefs, and is indeed grossly careless. One's carelessness is still in itself epistemically bad *pro tanto*. That is because it constitutes disrespect for the truth.[200] Maybe the consequences are also epistemically good—as opposed to simply good *and epistemic*—so that we ought epistemically *to cause ourselves* to think carelessly here. But it is a quite different question whether the carelessness would be in itself *pro tanto* epistemically bad. I think the answer to this question is "Obviously yes", just like in the case where disrespect for persons happened—thanks to some peculiar stage-setting—to promote aggregate well-being.

---

[199]Some views predict otherwise. But I think this is simply an objection to these views. Hurka, for instance, thinks that all pro-attitudes towards intrinsically valuable objects are themselves intrinsically valuable. Setting aside the objection I have to his misleading use of "intrinsic" discussed in the last note, I think he simply overgeneralizes from correct observations. What *is* true is that having the pro-response that is *fundamentally connected* with any value V is itself valuable, considered apart from its consequences. But we can accept this claim without accepting the far stronger claim that *any* correct pro-response to V is such that having that response to V will in itself be valuable, considered apart from its consequences. Hurka is guilty, I believe, of making this stronger claim on the basis of intuitions that only support the weaker claim.

[200]I take this type of case from Swank (2000: 201):

> Ours is in fact [such] a demonworld […] and the demon, Verity, has just let you in on the joke. She will presently put you back under, and you will […] forget all that she has just shown to you. But first she makes you an offer.
> As you are such a truth-lover, Verity can work some truth-conducive changes in your epistemic character. The new and "improved" you will be a paragon of [epistemic vice]—closed-minded, averse to argument and uncertainty, unappreciative of your own fallibility, and motivated by the desire to have been right. In consequence (this being [a] demonworld]) you will fare much better in the pursuit of truth than had you remained the […] person you are today. […] So: Would you accept Verity's offer? […] Neither would I.

Notice crucially that this is not just Cohen's "new evil demon problem". I here ignore the question of whether your beliefs would be epistemically justified in this world. Indeed, I ignore the question of whether the consequences of your carelessness would be epistemically good. I ask the simpler question: would your reckless disregard for truth in this world be a *pro tanto* epistemically bad thing in itself? The answer, I say, is "Clearly yes." This is the correct answer because truth is an epistemic value to be respected. If it were only an epistemic value to be promoted, there would be no reason to say this. For there is nothing *pro tanto* bad in *simply* not caring a whit about other values that are primarily to be promoted, like one's own pleasure. The pattern we here in the *evaluative* facts is not the one we expect on Weak Epistemic Teleology.

We should accept the simple intuition that respect for the truth is itself *pro tanto* epistemically good, irrespective of its consequences. This is *because* it is directed at something intrinsically epistemically good. This is independently intuitive, but also supported by our example. This casts direct doubt on Weak Epistemic Teleology, and supports the specific, epistemic Kantian reversal suggested in the last subsection. Again, Weak Epistemic Teleology leads us not to expect this pattern. Where V is a value primarily to be promoted, we can't reason as follows: "V is good. R constitutes love of or respect for V. So R is itself good, independently of its consequences." This reasoning only works when R is itself the fundamentally correct pro-response to V, which it will not be if V is a value to be promoted. Again: one's own pleasure is good, but patterns of motivation that express love or respect for one's own pleasure are not *themselves* good. That is exactly because one's own pleasure is primarily "to be promoted".

Notice that the Weak Epistemic Teleologist can't avoid the objectionable conclusion by holding that respect for the truth is just a further intrinsic epistemic value. This is because of the obvious asymmetry in value derivation: respect for the truth is in itself epistemically valuable *because* of the epistemic value *of that at which it is directed*. This asymmetry does not show it to be only instrumentally valuable. It rather shows, *contra* Extrinsic=Instrumental, that not all extrinsic value is instrumental value, at least if "extrinsic" means "derivative". Indeed, this is one of the counterexamples to that unfortunate assumption which underlies Weak Epistemic Teleology, and there are similar examples in the practical sphere. Respect for the truth *is* derivatively epistemically valuable in the sense of *deriving* its epistemic value from the epistemic value of truth (considered as a property of doxastic representations). It simply isn't instrumentally epistemically valuable. For that reason, we can continue to say truly that it is "in itself" good.

3.3.   *More support: a broader range of evaluative intuitions*

The last point leads to a stronger one. Weak Epistemic Teleology makes a mystery out of the *pro tanto* goodness of certain epistemic virtues that are not themselves truth-conducive—namely, certain motivational epistemic virtues. Epistemic Kantianism, on the other hand, directly explains this *pro tanto* goodness, and in a more elegant way I've seen on any other alternative to Weak Epistemic Teleology. Let me explain what I have in mind.

It is an epistemically *pro tanto* good thing when a subject's theoretical inquiry is driven by a disinterested respect for the truth. We can know that immediately, without taking a look at the results of that subject's inquiry. It is an epistemically *pro tanto* good thing *in itself*. Yet simply being driven by this motive obviously does not increase the objective probability one will believe correctly, or decrease the objective probability that one will believe incorrectly. It would be delusionally optimistic to think so. So, *pace* some theorists, it must be false that our intuition that this is an epistemic virtue is dependent on our prior belief that it will increase the objective probability that we will believe correctly if we have it. I certainly don't have such a delusionally optimistic belief. It is just clearly possible to identify doxastic conduct as *epistemically conscientious* regardless of what we know about the broader world in which it takes place. It is also possible rightly to judge that

epistemic conscientiousness is an epistemic virtue without such knowledge.

Some responsibilist virtue epistemologists have made this simple point over the years. Rightly, I say, though I am not myself a responsibilist virtue epistemologist. Montmarquet (2000: 139) gave a nice test case to illustrate this, though I'd dispute his diagnosis of the intuition's source:

> Suppose we were to discover that the world is actually a place in which the qualities—partiality, closed-mindedness, intellectual cowardice, etc.—we presently take to be epistemic vices are truth-conducive (and that the qualities we had previously taken to not to be truth-conducive were actually so). Now, in describing this *previous* state of affairs, what should we say? We could say that the qualities we previously took to be vices had all along been virtues (and that the qualities we previously took to be virtues were all along vices)—and that we had simply not known this. But such a proposal meets with this telling objection: that it severs the aforementioned ("Aristotelian") connection between the virtues and any reasonable notion of praise and blame. For surely we are not wanting to say that, all along, we should have been *blaming* people for being open-minded (and praising them for being closed-minded, etc.). Nor would it seem reasonable, beginning now, retrospectively to extend such judgments of blame and praise. […] We have, then a further reason to reject any quick and easy equation of the epistemic virtues with truth (or knowledge or understanding) conduciveness. For, obviously, what has changed is not whether these qualities are truth-conducive […].

A better way to put Montmarquet's point is as follows. Motivational epistemic vices like the ones he lists—e.g., partiality, closed-mindedness, intellectual cowardice, carelessness, etc.—are clearly *pro tanto* epistemically bad in themselves. Our inclination to call past instances of them epistemically bad would not change if we learned that these instances had, remarkably, been reliably truth-conducive. Similarly, motivational epistemic virtues like the ones he lists—e.g., impartiality, open-mindedness, intellectual courage, intellectual conscientiousness, etc.—are *pro tanto* epistemically good. Our inclination to call past instances of them good would not change if we learned that past instances of these traits had, remarkably, been systematically falsehood-conducive from the start. This is not plausibly explicable on Weak Epistemic Teleology. The most that the Weak Epistemic Teleologist can do is try to expand his list of intrinsic epistemic values. But this would be mistaken for reasons noted in previous subsections. It would get the *order of epistemic value derivation* wrong. It is clear that epistemic conscientiousness, which simply is a manifestation of respect for truth in belief, *derives* its value from the epistemic value of truth. It just derives it *noninstrumentally*, by manifesting the fundamentally correct response to this value.

To forestall objections, let me mark what I am *not* saying. I have only said that epistemic conscientiousness is necessarily *pro tanto* epistemically good in itself. Nothing I said implies that it wouldn't be epistemically *better* if some belief-forming policy were *both* (i) epistemically conscientious *and* (ii) truth-conducive. Epistemic Kantianism could be framed in a way that suggests this conclusion. After all, the demand to respect the truth requires us to constrain our deliberations by the rule to believe P only if P. To fail to constrain our deliberations by this rule would be to partly fail to have the relevant kind of respect for the truth. So *actually believing P only when P* by truly believing P in the particular case will often be *derivatively* epistemically good. It constitutes one kind of conformity with a rule by which we ought to constrain our deliberations because we ought more fundamentally to respect the truth. Epistemic Kantianism simply reverses the standard order of explanation. It is thus able to capture *more* than Weak Epistemic Teleology can capture. Because the epistemic value of truth is seen as *fundamentally* calling for respect, and motivational virtues of the kind we are here discussing simply partly constitute the relevant kind of respect, there is a direct explanation of why these virtues are *virtues*—i.e., *pro tanto* epistemically good traits for a believer to have.

It is also crucial to see that I am only assessing and making epistemically *evaluative* claims, not deontic or hypological ones. I am claiming that it is a necessary truth that certain motivational

epistemic character traits are *pro tanto* epistemically good in themselves. I framed all the relevant intuitions at the level of talk of epistemic virtues (or vices), which are by definition epistemically good (or bad) features for people to have. So, in making the claims I'm made, I was not begging the question against my earlier self, who encouraged us to see Cohen as being too hasty in claiming that the internal duplicate of someone epistemically justified in believing P in the actual world would be equally justified in believing P in the demon world. Cohen was too hasty. Nevertheless, what we see is that there are some related points that survive our critique of him.

3.4.   *More support: when true beliefs aren't epistemically good and false beliefs aren't epistemically bad*

Another reason to prefer my specific reversal is that it predicts plausible asymmetries in when truly believing is *pro tanto* epistemically good and falsely believing is *pro tanto* epistemically bad that we would not otherwise expect. As we will see in the next subsection, it predicts a relatedly plausible asymmetry in when we ought epistemically to promote the Conditional T-Goal. This, we'll see, will lead to the rejection of Strong and not just Weak Epistemic Teleology.

To bring out the first asymmetry, let's reflect on how the epistemic Kantian picture is entitled to regard some true beliefs as epistemically good and some false beliefs as epistemically bad. Once again, the core claim of the epistemic Kantian conception of epistemic value is that truth is an epistemic value primarily to be respected. Being more specific than before, it makes this claim:

> (**Kantian Conception of Truth's Epistemic Value**)
> Truth (viewed as a property of beliefs) is an epistemic value to which the fundamentally correct response is respect as a standard (in the particular case), and to which the fundamentally incorrect response is disrespect as a standard (in the particular case). To properly value truth in forming or holding some doxastic attitude *vis-à-vis* P is thus to be fully successfully constrained in one's reasoning with respect to P by the injunction *believe P only if P*, since that is just what it is to fully respect truth as a standard with respect to P.

As I said before, the epistemic Kantian agrees that we ought to constrain our doxastic deliberation in the particular case by the injunction expressed by the Conditional T-Goal. After all, to fail to be deliberatively constrained by this injunction trivially entails failing to be fully constrained by the injunction to believe P only if P. On the formulation of the Kantian conception just given, that would *ipso facto* be to fail to respect truth as a standard for belief in the particular case. Given the epistemic Kantian conception, that would be to fail to correctly value truth, understood as a property of doxastic representations. And that is epistemically bad.

But how does the epistemic Kantian derive conclusions about the epistemic goodness or badness of *particular cognitive states* or *intellectual acts*? She needs some account of epistemic value derivation. She should, I believe, agree to the following highly general, neutral account:

> (**Response Derivation**)   Some cognitive state or intellectual act would be
>
> > (i)   extrinsically epistemically (*pro tanto*) good if it manifests a *fundamentally correct* response to intrinsic epistemic value,
> >
> > and
> >
> > (ii)   extrinsically epistemically (*pro tanto*) bad if it manifests a *fundamentally incorrect* response to intrinsic epistemic value.

Notice that Response Derivation is a constraint on value derivation that even the Weak Epistemic Teleologist accepts. The Weak Epistemic Teleologist thinks that the fundamentally correct response to epistemic value is to *bring it about*, and the fundamentally correct response to epistemic disvalue is to *eliminate it*. Coupled with Weak Epistemic Teleology, Response Derivation suggests

that extrinsic epistemic value will be instrumental value, or value as a means to ultimate epistemic goals. That is exactly what the Weak Epistemic Teleologist will want—if she is honest, at any rate. The epistemic Kantian agrees about Response Derivation as a general account of epistemic value derivation but rejects the Weak Epistemic Teleologist's picture of the fundamentally correct response to intrinsic epistemic value. So she instead upholds:

> (**Epistemic Kantian Derivation**) Some cognitive state or intellectual act directed at a proposition P would be extrinsically epistemically (*pro tanto*) good to the extent that it manifests respect for the truth with respect to P in the sense given by the Kantian Conception, and extrinsically epistemically bad (*pro tanto*) to the extent that it manifests disrespect for the truth with respect to P in the sense given by the Kantian Conception.

Epistemic Kantian Derivation makes some important predictions. It predicts

> (a) that true belief in P will not be *pro tanto* epistemically valuable insofar as it fails to manifest respect for the truth *vis-à-vis* P,

and

> (b) that false belief in P will *not necessarily* be *pro tanto* epistemically disvaluable if it doesn't manifest disrespect for the truth *vis-à-vis* P.

These predictions are, I believe, plausible ones. Suppose some epistemic subject S altogether lacks apparent sufficient reasons to believe P, and indeed has apparent sufficient reasons to doubt that P is true. Suppose that she believes that P in the face of these apparent sufficient reasons. Remarkably, by the lights of the perspective on the facts available to her, P turns out to be true. Is there anything *pro tanto* epistemically good in her believing truly here? I think it is more intuitive that the answer to this question is "No" than that it is "Yes". True believing isn't *ipso facto* better believing, where that is read attributively. When true believing would also be *totally careless believing*, as it is here, it simply isn't good believing in the relevant sense. The carelessness *undercuts* the epistemic goodness of believing truly. Epistemic Kantianism correctly predicts this. Weak Epistemic Teleology entirely lacks the resources to predict this in a way that would also get intuitive patterns of value derivation correct, as I argued in the last subsection.

To the extent that this does not sound like an obviously correct conclusion to you, you are, I think, confusing what is *good and epistemic* with what is *epistemically good*. I certainly agree that true beliefs are often *pro tanto* good period. True beliefs are very useful. They get us to Larissa. Moreover, there is something plausibly valuable *simpliciter* in accurately rather than inaccurately representing reality. Alston got all of this right in the passage we quoted much earlier. Since true beliefs are epistemic items in some sense, I also agree that true beliefs are always *pro tanto* good and epistemic. But it simply does not follow that they are epistemically good. For my own part, once I distinguish these two questions, it is highly intuitively appealing to me that believing on a careless, reflectively lucky guess is *not good believing* at all, in the relevant attributive sense.

I should emphasize that I am not denying that beliefs are in some sense "correct" iff true! Understood *truly*, this claim is simply orthogonal to the discussion. Thomson put the point well:

> [S]uppose you came believe that P, and therefore asserted that P. And suppose it turned out to have been the case that P. It is not at all plausible to think that your believing [*ipso facto*] turned out to have been better *qua* believing than it otherwise would have been […].
>
> Similarly for the perfect nominals. For a belief to be a correct belief is just for it to be a true belief, and a true belief is not better *qua* belief than a false one—as for an assertion to be a correct assertion is just for it to be a true assertion, and a true assertion is not better *qua* assertion than a false one.

> I invite the conclusion that 'being a correct believing' and 'being a correct belief' are not favorable evaluative properties. [*Nota bene*: Thomson means "attributive evaluative properties", since for her all evaluative properties are attributive and none are predicative.][201]

Now, I don't agree with Thomson's final claim when read in an *unrestricted* way. Truth in belief is a good thing *simpliciter*, as I've said. And I doubt we should agree with Thomson that all evaluation is fundamentally attributive. But I agree with Thomson that it is not *by itself* an epistemically good thing *simply* to believe correctly, where "correctly" is understood so that "belief is correct iff true" is plausible. This is because it is not automatically better *qua* believing. Cases like the one I just discussed should make this vivid and plausible, at least to minds uncontaminated by the historical conflation of "good and epistemic" with "epistemically good". The plausibility of this claim should lead us—as it led Thomson—to deny that correct believing in the relevant sense is by itself good believing *qua* believing. Of course, given that we live in a friendly world, correct believing is often better *qua* believing than incorrect believing. Methods that guarantee respect for the truth will often be reliable in this comparatively friendly world, and *vice versa*. But we shouldn't be misled by this. Thankfully, simple reflection on particular cases discourages confusion here—as epistemic Kantianism predicts it should.

Of course, it is not altogether obvious what it even is for believing to be good *qua* believing. Given the plausibility of its predictions and our other arguments for it, I suggest that we accept the epistemic Kantian picture: believing is (I) best *qua* believing when it manifests full respect for the truth with respect to the question that the belief is aiming to answer, (II) good *qua* believing when it manifests sufficient respect for the truth in this proposition-relative way, and (III) bad *qua* believing when it manifests disrespect for the truth in this proposition-relative way.

3.5. *More support: asymmetries in promoting the Conditional T-Goal*

How, you might ask, is this conclusion compatible with what I said in §3.1? Did I not say there that the epistemic Kantian thinks we should be deliberatively constrained by the Conditional T-Goal? Yet, in point of trivially obvious necessity, false beliefs can't promote the Conditional T-Goal. Aren't these claims inconsistent with the claims I made at the end of the last subsection?

They are not. The epistemic Kantian thinks that we ought *to successfully constrain our theoretical deliberation* by the injunction expressed by the Conditional T-Goal, since doing this is partly constitutive of having the relevant kind of respect for the truth. Yet from the fact that

> we ought to constrain our deliberations by a rule R,

it does not follow that

> we ought in fact to objectively conform to R in every case.

This inference fails when the fundamental reason why we ought to constrain our deliberations by some rule R is because doing so is constitutive of respecting a certain value addressed by R. In these cases, if clear apparently good total evidence misleadingly suggests that we would violate R by A-ing but conform with R by B-ing, respect for the value in question will demand B-ing rather than A-ing *even if* B-ing would *not* yield conformity with R and A-ing would. For once again, respect directly requires constraining one's deliberation by R. It does not directly require conforming to R. These things can come apart.

---

[201]Thomson (2008: 112).

As an illustration, we can consider the following variation on *Ingenious Marital Reconnaissance*.

> *Doubly Ingenious Marital Reconnaissance*. Jane worries that her husband Dale would cheat on her if he had the opportunity. She decides to test this hypothesis in a clever way. Using her extraordinary costuming skills, she manages to dress up to look like a different woman on whom she suspects Dale would have an instant crush. Since strongly expects her hypothesis to be confirmed, she decides to take over this role entirely for a while, just for droll fun. So she also pays one of her friends to be disguised to look exactly like her, and live in her home with Dale for the next few weeks. Dale is none the wiser. He is also careless. He starts flirting and making more ambitious arrangements with the woman in disguise whom he meets outside his workplace, becomes cold and unaffectionate toward the woman he believes to be his wife, and frequently returns home late and sleeps in his office, pretending to work.

Just as in the original *Ingenious Martial Reconnaissance* case discussed at the end of Chapter 3, Dale plausibly lacks the kind of respect he *really ought* to have for his wife. She has cause for complaint. Given that this is what matters for their relationship, he does what he really oughtn't do. This is in the genuinely normative sense of 'oughtn't' that we care about. Plausibly, part of the reason why is that he fails to constrain his practical deliberation by the following directive:

> R: Don't have an extramarital affair and stop giving your wife the affection she deserves.

What is the alternative? If Dale tries to comply with R, he will not conform to R. He will end up romancing someone other than his wife, perhaps with considerable passion! So: (i) Dale ought to constrain his practical reasoning by R, because doing so is necessary for having the kind of respect he ought to have for his wife, and that is what is most important to the health of their relationship in this case, and therefore (ii) he ought not actually to conform to R.

Once again, I don't think our intuitions are simply conflating the 'ought' of rationality with the 'ought' of reasons in this case. What there is most reason for Dale to do in this case is respect his wife. That is what is central to the value of his relationship with her. To do that, he must govern his deliberations by R, and so do what would ironically fail to constitute conformity to R in this case. If epistemic Kantianism is true, a parallel conclusion holds for us. To respect truth as the standard for belief formation, one must constrain one's deliberations by the injunction expressed by the Conditional T-Goal. Since respect is the response fundamentally called for by the epistemic value of truth on this view, one ought sometimes not to conform to the injunction. This will be true when conforming to it would require rashly disrespecting the truth, and so carelessly failing to have one's doxastic deliberation constrained by the injunction.

How can I illustrate this? Well, there are many cases where the epistemic Kantian prediction is *prima facie* plausible. Of course, however, many of these cases are ones where untutored intuition *will* conflate epistemic rationality with epistemic justification. And I have urged us in previous chapters to avoid this conflation. Nevertheless, as we'll see in Chapter 6, while it is true that rationality and justification are as *notionally distinct* in epistemology as they are in the practical sphere, they are extensionally—and intensionally—much closer. I will, indeed, *use* epistemic Kantianism to argue for this. But at present, invoking this conclusion would beg the question. So for now I'll continue to set these cases—e.g., demon world cases as in Cohen (1984)—aside.

There are, however, two less controversial types of case we can consider. They are cases where Strong Epistemic Teleology has a difficult time yielding the right predictions. I discussed some of these cases at the end of Chapter 3 in noting how the problems of deontic significance for epistemic rationality matter for many standard externalist theories of *ultima facie* justification as well as internalist theories of *prima facie* justification. As I noted, it is much harder to accommodate

simple intuitions about defeat than many externalists suppose. After all, surely

> a *prima facie* bad reason against believing P cannot make a *prima facie* good for believing P a worse candidate for being an *ultima facie* good reason for believing P.

But let the criterion of quality for epistemic reasons be the objective teleological one expressed by the Reasons Quality Formulation of Strong Epistemic Teleology. This criterion together with the principle just expressed suggests that if an (indicator) unreliable *ultima facie* nondoxastic experience as of ~P arises after one has gained *prima facie* justification for believing P, one's *prima facie* justification for P is not thereby outweighed. More strikingly, the obvious verdicts about clairvoyance cases become hard to license if we cling to strongly teleological criteria of quality. If clairvoyance experiences can be reliable indicators of the facts, it is unclear why they can't provide reasons for belief as good as ordinary perceptual experiences—at least given the criteria of quality that explained why the ordinary perceptual experiences can provide good reasons.

Epistemic Kantianism, on the other hand, predicts an asymmetry in when we are epistemically permitted to promote the Conditional T-Goal. If actually promoting the Conditional T-Goal in a given case would be incompatible with respect for the truth, we ought not to do what would guarantee conformity to the Conditional T-Goal. Again, this is not only consistent with the view— as I argued earlier in this section—but also directly predicted by it.

This, I suggest, is what is happening in these cases. In cases where ultimate nondoxastic perceptual experience suggests that ~P *after* one has gained *prima facie* justification for believing P, one is no longer permitted to believe P. Why? Because doing so would be obviously inconsistent with respect for the truth about whether P. That is regardless of the unreliability of this experience type in indicating the facts. Moreover, in cases where one is struck with a new kind of *prima facie* nondoxastic seeming as of P but lacks ultimate apparent reasons to think that this *prima facie* nondoxastic seeming is issuing from a source that yields reliable indications of the facts, one is not permitted to believe that P. Why? Because doing so would be obviously inconsistent with respect for the truth. The prediction made by epistemic Kantianism can thus be used to underpin the most plausible verdicts about the defeat of justification. As I insisted in the last chapter, although many externalists (e.g., Goldman (1986)) have constructed accounts of *ultima facie* justification that also make these predictions, it is unclear how the standards of quality for epistemic reasons on which these theories must fundamentally rest can really vindicate these predictions. The clauses for defeat that get tacked on in these theories really should look *ad hoc* and indeed implausible by the lights of the guiding criteria of quality for reasons that motivate their initial accounts of *prima facie* justification.

Given that we have other reasons to reject Weak Epistemic Teleology—as I argued in the last few subsections—we can doubt whether those standards of quality are correct. We can see epistemic Kantianism as offering the only clear foundational explanation of how these defeat phenomena are possible. It is not enough to note that there are other theories that make these predictions. There must be a coherent foundational motivation for these theories—one that is not *ad hoc* by the lights of the criteria of quality that get them off the ground. A theory should not simply *cherrypick* the intuitions it is going to honor. It must have a unitary basis for honoring some intuitions and dismissing others, and for tacking on certain clauses to handle counterexamples and avoiding other clauses. Epistemic Kantianism is unique in its ability to avoid cherrypicking here—at least in comparison with a great many theories I've addressed.

3.6.   *Taking stock: why Weak Epistemic Teleology fails*

Since the reasoning has been fairly intricate, let's take stock of what has been achieved.

In this section and the last, I've been arguing against an axiological form of epistemic teleology. It is a view more often presupposed than explicitly defended that I call Weak Epistemic Teleology. Weak Epistemic Teleology claims that what it is to be epistemically valuable is to be a goal that we ought epistemically to achieve (intrinsic case) or to be instrumentally conducive to such a goal (extrinsic case), and that what it is to be epistemically disvaluable is to be an "anti-goal" that we ought to avoid promoting and avoid (intrinsic case), or to be instrumentally conducive to such an anti-goal (extrinsic case). My reason for attacking this lesser discussed view is that if it fails, Strong Epistemic Teleology fails—or so I'll argue in the next section.

The overall argument can be broken into three stages. The first stage was to point out that this view would only seem plausible if one made certain axiological assumptions that are widely discredited in the broader value-theoretic literature, and also to note that there is little reason to expect the domain of epistemic value to be any different once we've distinguished between (i) values and goals that are epistemic and (ii) properly epistemic values and goals. Chief among these axiological assumptions are the assumption (a) that intrinsic value is final value—i.e., value to be brought about as an end in itself—and (b) that extrinsic value is instrumental value.

Rather than just listing counterexamples from the broader literature, I gave a theoretical argument against (a) and (b). It came from the connection between *values* and *fundamentally correct or fitting ways of valuing*—a connection that a majority in the broader value-theoretic literature, buck-passers and buck-stoppers alike, accept. Given the connection, the argument was:

(1) Intrinsic=Final and Extrinsic=Instrumental are plausible only if

(a) the only fundamentally correct response to intrinsic value is to bring it about for its own sake as an end,

and

(b) the only fundamentally correct response to extrinsic value is to bring it about as a means to such an end.

(2) There is no reason to believe (a) and (b). Indeed, there are clear counterexamples. Consider the intrinsic value of persons. Narveson's asymmetry—i.e., that we are in favor of making people happy but neutral about making happy people—undercuts the plausibility of (a) in this case. The best explanation of Narveson's asymmetry—viz., that the value of persons is fundamentally linked with *respect* and not *production*—directly rebuts (a). Once again, one cannot respond by saying that what's *really* intrinsically valuable are just *acts of respect* or the *conditional state of affairs of people-being-happy-if-they-exist*. Acts of respect and this conditional state of affairs have *extrinsic* value—value that is *derived from* the intrinsic value of persons. If persons weren't valuable, respecting them wouldn't be valuable and the conditional state of affairs wouldn't be either. It is just that this extrinsic value isn't instrumental. So lurking in this connection is a counterexample to (b) as well as (a).

(3) So, Intrinsic=Final and Extrinsic=Instrumental should be rejected.

The reason why I focused on this argument rather than the hosts of particular counterexamples that people have leveled against Intrinsic=Final and Extrinsic=Instrumental is that it brings out *how* Weak Epistemic Teleology can fail if these prior assumptions fail.

This thought brought me to the second stage of my argument. As I noted at the outset in this section, there is a structural analogue of Narveson's observation in epistemology:

(**Analogous Asymmetry**) While we ought epistemically to have the aim of seeing to it that

our beliefs are true, we have no non-optional epistemic reasons
to have the aim of producing more true beliefs.

Appealing to practical restrictions to explain the second half of the Analogous Asymmetry was dismissed as conflating epistemic goals with goals *simpliciter* that are epistemic. The Weak Epistemic Teleologist ought, I said, to be puzzled by this asymmetry. If true belief is an intrinsic epistemic value and Weak Epistemic Teleology is true, we should *not expect* the Analogous Asymmetry. More strongly, since truth (in belief) clearly is an intrinsic epistemic value and this asymmetry exists, we should doubt Weak Epistemic Teleology. For the *best explanation* of this asymmetry is precisely like a standard explanation of Narveson's asymmetry: truth (in belief) is an epistemic value to which the fundamentally correct response is *respect*, and *not production*.

The Weak Epistemic Teleologist cannot appeal to this explanation. He must take the Analogous Asymmetry to be a *brute (normative) fact*, and adopt a bizarre picture of what is actually intrinsically epistemically valuable. He must hold that instances of the *conditional state of affairs type* of a *belief-being-true-if-we-have-it* is what is really intrinsically epistemically valuable. For only that is plausibly a fundamental truth-connected epistemic goal of ours, given the Analogous Asymmetry. This is an unappealing position. It is structurally like a defender of Intrinsic=Final saying that what is really intrinsically valuable is the conditional state of affairs of *people-being-happy-if-they-exist*. Neither view should be accepted, and we should look for an explanation of *why* these asymmetries exist. The best explanation is, as I've suggested, structurally the same.

Having made these points, I turned to the final stage of my overall argument, which involved providing further support for the particular reversal of the order of epistemic value derivation that I used to explain the Analogous Asymmetry. I pointed to four sources of support:

> (I)   It gets some patterns of epistemic value derivation right that Weak Epistemic Teleology (and other plausible alternative views) cannot get right. It correctly predicts that respect for the truth is *in itself* epistemically good and yet is also epistemically good *because* it is respect for something that is itself intrinsically epistemically valuable—namely, truth (understood as a property of beliefs). In other words, it correctly predicts that respect for the truth is epistemically good in a noninstrumental but extrinsic way that derives more from the epistemic value of truth (understood as a property of beliefs).[202]

> (II)   It explains why certain cognitive motivational virtues are by themselves necessarily *pro tanto* epistemically good in themselves, and why certain cognitive motivational vices are by themselves necessarily *pro tanto* epistemically bad, regardless of their consequences.

> (III)   It explains why truly believing is not always *pro tanto* epistemically good, and why falsely believing is not always *pro tanto* epistemically bad—a conclusion that is appealing once we distinguish what is good *simpliciter* and epistemic from what is epistemically good.

> (IV)   It relatedly explains why there is an asymmetry in when we are permitted to promote the Conditional T-Goal—specifically, why we are not epistemically permitted to promote this goal if promoting it would require disrespecting the truth with respect to some proposition.

I'm sure there are other virtuous predictions worth noting, but I'll leave things here for now.

It is now time to begin reaping the payoffs of the rejection of Weak Epistemic Teleology. The

---

[202]Here and elsewhere "in itself" does not mean "intrinsically". I understand "intrinsically" as "nonderivatively". Something can be good *considered on its own*—e.g., love of something intrinsically valuable—but still *derive* value from one of its components (here, the intentional object of the particular love).

fountainhead of payoffs is that if Weak Epistemic Teleology fails for the reasons I've suggested, Strong Epistemic Teleology also fails. This is a payoff because it will enable us in the next two chapters to see why arguments against the deontic significance of epistemic rationality fail, and also to construct a positive explanation along epistemic Kantian lines of *why* epistemic rationality has deontic significance *via* a more appealing picture of the place of truth in epistemology. Before jumping to these upshots, I'll turn to explain how Strong Epistemic Teleology fails.

4.   A Scanlonian Argument against Strong Epistemic Teleology

4.1.   *A note on the inspiration for the argument*

To bring out the structure of my argument, it may help to say a bit about its inspiration. It is inspired by what I take to be the best way of opposing act consequentialism in ethics.

Act consequentialism has most often been framed as the view that what is right is just what promotes the most value, where value is taken to be understood antecedently to the right.[203] Opposition to act consequentialism has most often appealed to intuitions about particular cases that seem to show that there are *side-constraints* or *restrictions* on the promotion of value. This familiar type of opposition to act consequentialism *concedes* that to be valuable is fundamentally "to be 'to be promoted'", and is intelligible antecedently to the right. All of our *value-based* reasons for action are *granted* by the opposition to be reasons that promote the good. The opposition simply says that not all of our reasons for action are value-based. Some of our duties are *just plain duties*: we know that they apply with great force prior to knowing what amount of value would be brought about by acting in accord with them. Sometimes we must fulfill these duties even when doing so brings about a substantially worse state of affairs. We are supposed to see this by reacting to many now familiar cases—e.g., the standard array of "trolley" cases.

This is not, to my mind, either the deepest or the most satisfying way to oppose act consequentialism. It is better to refuse to grant to the act consequentialist his assumptions about the nature of the good—specifically, his assumption that to be good or valuable is, as Scanlon nicely put it, "to be 'to be promoted'". One reason why this style of opposition is better is that there *is* something very appealing about the simple idea that we ought to do the best thing, and that the quality and strength of our reasons for action are fixed by the good in successfully responding to them. Consequentialists slide from these simple, appealing ideas to an idea that can *sound* equally appealing: namely, that we ought to do what would *make things go best*. It will not, I believe, be illuminating to deny that this claim is true *just* by appealing to intuitions about cases to support restrictions—intuitions which are not clearly reliable, and which can also often be explained away. Indeed, the appeal of the much weaker idea that we are *permitted* to bring about the best outcome will put pressure on quick retreats to restrictions. A satisfying critique of act consequentialism will be one that can *explain why* the move from "Do the best thing" to "Bring about the best outcome" is invalid, and explain why the intuition that we ought to do the best thing *does little or nothing* to decide between extensionally divergent half-plausible theories.

So I think the way to run the critique is to attack the core idea that "to be valuable is to be 'to be promoted'". If correctly responding to value does not fundamentally involve bringing as much

---

[203]Strictly speaking, this is only the claim of *maximizing* act consequentialists. *Satisficing* act consequentialists will claim instead that what is right is just what promotes "enough" value. They typically do this to avoid complaints about over-demandingness. Satisficing is not, I believe, a wise move. It is far better to explain away the intuitions about demandingness by drawing a deontic/hypological distinction. We ought to do the best thing, but when that would be particularly hard, we can be partially or fully excused for acting wrongly. But regardless of whether I'm right about that, the satisficing/maximizing distinction will have little importance for the points that follow. So I'll focus on the maximizing view for simplicity's sake.

of it as we can (or "enough" of it), we can understand both (i) how we can be required to do what's best, and (ii) how we can be permitted to do what wouldn't bring about the *state of affairs* that contains the greatest (or nearly the greatest) *quantity of good*. This, among other things, is what Scanlon (1998) did. And with great appeal. He argued plausibly that our ordinary thinking about value cannot be cast as the consequentialist presupposes. The spirit of Scanlon's attack was, I agree, on target. I remain attracted to consequentialism *only* because I think our reasons to make things go better for people are the strongest, and outweigh the reasons we have to respond to value in non-teleological ways. I agree with Philippa Foot in thinking that beneficence isn't the only virtue. I simply continue to think it is decidedly the most important of the virtues.

I won't, however, try to get into why I think *this* here. What matters here is that my reasons are not ones that have analogues in the case that interests us: mine are substantive reasons having to do with the nature of certain practical values and their relative importance. Predictably, then, I'm optimistic that a successful attack on Strong Epistemic Teleology exists that has the form of Scanlon's attack. I have already prepared the grounds by attacking Weak Epistemic Teleology. This is the epistemic analogue of the simpler view Scanlon attacked—viz., that to be valuable is "to be 'to be promoted'". I'll now extend the Scanlonian argument to epistemology.

4.2.   *The Scanlonian argument against Strong Epistemic Teleology*

The argument begins with a point of agreement with the Strong Epistemic Teleologist. The Strong Epistemic Teleologist is naturally understood as presupposing this

> **Value-Driven Constraint on the Criteria of Quality for Reasons**
> R is a *prima facie* good epistemic reason for believing P *only if*, setting other epistemic reasons bearing on whether to believe P aside, believing P for R makes it objectively (more) likely that one will correctly respond to intrinsic epistemic value simply in believing P.

She is then naturally seen as deriving the Reasons Quality Formulation of Strong Epistemic Teleology I formulated at the very beginning of this chapter from Weak Epistemic Teleology and the Value-Driven Constraint. Specifically, she grants herself the following

> **Implication of Weak Epistemic Teleology**
> To correctly respond to intrinsic epistemic value simply in believing P *just is* to promote the core epistemic goals with respect to P in simply believing P.

This is an implication of Weak Epistemic Teleology because, as I argued earlier, that view requires that the fundamentally correct response to epistemic value just is to promote it either as an ultimate goal or as a means to one. If the fundamentally correct response to epistemic value is just to promote it as a goal, then, of course, to correctly respond to epistemic value *simply in φ-ing* would be to promote core epistemic goals simply in φ-ing. Hence this implication.

The Implication of Weak Epistemic Teleology together with the Value-Driven Constraint directly entail the Reasons Quality Formulation of Strong Epistemic Teleology. To satisfactorily oppose the Strong Epistemic Teleologist, it is advisable to reject one of these assumptions.

But the Implication of Weak Epistemic Teleology is, as we saw in the last two sections, false. This is because Weak Epistemic Teleology is itself false. Once we've given up Weak Epistemic Teleology, we can *grant* the Value-Driven Constraint and simply uphold the

> **Non-Teleological View of Correctly Responding to Intrinsic Epistemic Value**
> To correctly respond to intrinsic epistemic value simply in believing P is to respect the truth

about whether P in believing P, and this is often (modally speaking, at any rate) not to promote the most plausible epistemic goal (i.e., the Conditional T-Goal).

This Non-Teleological View—which is just the epistemic Kantian one we've provisionally motivated in the last two sections—entails the falsity of Strong Epistemic Teleology, at least if we accept the Value-Driven Constraint that the Strong Epistemic Teleologist accepts.

So, Strong Epistemic Teleology fails, given that the Value-Driven Constraint holds and Weak Epistemic Teleology fails.

Let's put the argument more officially:

The Scanlonian Argument

1.   The Value-Driven Constraint holds. Strong Epistemic Teleology gets this right.

2.   But Weak Epistemic Teleology is false.

3.   If Weak Epistemic Teleology is false, the best replacement for it will be the Non-Teleological View just stated.

4.   But the Value-Driven Constraint and the Non-Teleological View entail that Strong Epistemic Teleology is false.

5.   So, Strong Epistemic Teleology is false.

We defended (2) and (3) in the last two sections. (4) is made plausible by the point I made in §3.5. There, I noted that although the most plausible non-teleological view will require that we *constrain our deliberations* by the injunction expressed in the Conditional T-Goal, it does *not* require that we *actually conform* to this injunction. Indeed, it may require that we *not* conform to this injunction, if conforming to it with respect to some proposition P would require failing to respect the truth with respect to P. If that is right, the non-teleological view will imply that correctly responding to the epistemic value of truth can require *not doing* what would make it objectively most likely that we would promote the only properly epistemic goal with respect to P in believing P. That is just to say that Strong Epistemic Teleology is false, since Strong Epistemic Teleology will always require doing what would be most likely to promote the core epistemic goals with respect to P in believing P. So, granting to the Strong Epistemic Teleologist what he himself needs to get his view running, we can refute his view.

4.3.   *Can the Strong Epistemic Teleologist subsume our view's implications by retreating from veritism?*

Now, there is an important kind of response worth considering on behalf of the Strong Epistemic Teleologist. I have been focusing on *veritist* versions of Strong Epistemic Teleology. I said at the outset of the chapter that this would make no difference. One might, however, wonder whether that is true, and reasonably demand me to explain why I am confident that it is.

The possibility of a non-veritist view does raise the following structural worry about the Scanlonian Argument as I've rendered it. If we aren't veritists, couldn't fulfilling the core epistemic goals with respect to P simply *require respecting the truth* with respect to P? If it could, the resulting non-veritist version of Strong Epistemic Teleology *would have the same good predictions* that the epistemic Kantian view has. If so, then (4) in the argument will not be true. Indeed, this point *does* show how the argument that I gave for (4) at the end of the last subsection could fail, since that argument was addressed specifically to the *veritist* teleologist who agrees that the fundamental

epistemic goal is something like the T-Goal, conditional or unconditional.

Still, I don't think there is a plausible response along these lines. To see why, let's first narrow down the range of options. Nobody will think that *simply being epistemically rational* is itself a fundamental epistemic goal, or an intrinsic epistemic value. Even the epistemic Kantian does not think this is true. The epistemic Kantian thinks that, in theoretical deliberation directed at answering the question whether P, we ought to respect the truth with respect to P. This does, as I've said in passing and will argue in the next two chapters, require us to be epistemically rational. Complying with requirements of rationality *is* the way to respect the truth with respect to P. Still, the reason to be epistemically rational on this view is *derived*: we ought to be epistemically rational, on this view, *because* doing so is necessary for respecting the truth, and the fundamentally correct pro-response to the epistemic value of truth is to respect it. So I think we can safely ignore views that would make being epistemically rational itself one of the *core* or *fundamental* epistemic goals, or make epistemic rationality one of the intrinsic epistemic values. It is true that if the Strong Epistemic Teleologist of the stripe I've been discussing in this chapter accepted this view, he could block my argument by rejecting (4). But this is not the way in which he, or anyone, will conceive of the place of epistemic rationality in epistemology.

So retreating from veritism to a view that puts epistemic rationality first isn't going to work. There are, however, a couple of other candidate views. In the recent literature on epistemic value, some people have been attracted to the following views:

> **The Knowledge View**: Knowledge is an intrinsic epistemic value (and a fundamental epistemic goal).
>
> **The Understanding View**: Understanding is an intrinsic epistemic value (and a fundamental epistemic goal).

Could these views help the Strong Epistemic Teleologist avoid the Scanlonian argument? I don't think so. There are two reasons. One reason is that the resulting versions of Strong Epistemic Teleology of the type I've been considering would not be plausible. Take the Knowledge View first. Accepting the Knowledge View would recommend the following minimal variation on the Reasons Quality Formulation of Strong Epistemic Teleology:

> **Knowledge-First Strong Epistemic Teleology** (*Reasons Quality Formulation*):
>
> R is a *prima facie* good epistemic reason for believing P *only if*, setting other epistemic reasons bearing on whether to believe P aside, believing P for R makes it objectively (more) likely that one will know the truth with respect to P.

Here is an obvious objection to this view. Suppose one has a visual experience as of P, and that visual experiences are in fact highly reliable indicators of the truth. Suppose moreover that one's visual experience happens to be, unbeknownst to one, a veridical hallucination. Believing P for the reason given by this experience will *not* make it objectively more likely that one will know. Indeed, it will guarantee that one will *not* know, since if one only believes something for the reason given by a veridical hallucination, one will be Gettiered and hence not know. Veridical hallucinations produce the kind of luck or accidentality that excludes knowledge. So this Reasons Quality Formulation of Strong Epistemic Teleology will imply that the reason given by your veridical hallucination is not even a *prima facie* good reason. This is clearly wrong. Even if it is not a *sufficient* reason for belief, surely it is a *prima facie* good reason. This view holds the quality of epistemic reasons hostage to worldly contingency in a way that simply cannot be plausible.

There may be ways of revising the view to avoid this implication. But there is a much simpler

problem that arises from points I made in Chapter 3. Recall that I argued in Chapter 3 that rationality can require giving up knowledge. This is true in some forgotten evidence cases. Since respect for the truth requires complying with the requirements of rationality, it follows that respect for the truth can require giving up knowledge. Yet remember what the strategy for blocking the Scanlonian Argument was going to be. The idea was to argue from this premise:

fulfilling core epistemic goals with respect to P requires respecting the truth on whether P.

It is true that if one could argue for that premise, the Scanlonian Argument could be blocked. Yet the Knowledge View's picture of our core epistemic goals will *not* enable one to make this response for the reason just noted. Rationality can require giving up knowledge with respect to P. So can respect for the truth about P. So respect for the truth about whether P can require not fulfilling the "knowledge goal" with respect to P. If so, retreating to the Knowledge View *won't allow one* to pursue the strategy needed to block the Scanlonian Argument.

So this retreat is a nonstarter. The same problem arises for the Understanding View. But the more important thing to note is that no Understanding View could *possibly* give one an adequate account of the criteria of quality for epistemic reasons. Here is a simple argument:

1.   We can have good reasons to believe propositions that do not contribute to our understanding of anything (e.g., propositions about the numbers in a telephone book, say).

2.   If (1), then the criterion of quality for these reasons cannot be measured by the extent to which complying with these reasons would (be likely to) increase one's understanding with respect to the propositions that they are reasons for.

3.   So the criterion of quality for these reasons cannot be measured by the extent to which complying with these reasons would (be likely to) increase one's understanding with respect to the propositions that they are reasons for.

4.   But if (3), then there could not be a version of Strong Epistemic Teleology designed to give criteria of quality for reasons that relevantly profits from the putative fact that understanding is a fundamental epistemic goal (or one of the intrinsic epistemic values).

5.   So there could not be a version of Strong Epistemic Teleology designed to give criteria of quality for reasons that relevantly profits from the fact that understanding is one of our fundamental epistemic goals (or one of the intrinsic epistemic values).

6.   That's just to say that the Understanding View can't block the Scanlonian Argument.

The only response I can see is to reject (1), and argue for a very weak picture of understanding on which truly believing anything contributes to our understanding. But if one retreats to such a picture of understanding, it will become extremely implausible that such understanding really is one of our core epistemic goals, or an intrinsic epistemic value "to be promoted". So there is actually a dilemma. If understanding is understood strongly, the argument just given goes through. If it is understood weakly, the Understanding View is false, and so cannot help.

I'm not aware of other obvious candidates for what the intrinsic epistemic value or ultimate goal might be if it isn't truth. Knowledge and understanding are the most natural candidates. They are the ones that most often get discussed as alternatives in the literature. Respect for the truth, as I've insisted, is not itself an intrinsic value. Respect for the truth derives its epistemic goodness from the fact that it is the correct response to the epistemic value of truth. That assumption should be preserved on any view. It is highly plausible. So the epistemic Kantian view does not

put respect first among intrinsic epistemic values or ultimate epistemic goals. It views respect as the fundamentally correct response to the epistemic value of truth. It is not itself usefully understood as a goal. It is just the response required by the intrinsic epistemic value of truth. So retreating to a view that would simply translate the epistemic Kantian's claims into claims about what the intrinsic epistemic values or final goals are would not help. It would, indeed, simply contravene the driving motivations for accepting the view in the first place.

So I do not see a response to the Scanlonian Argument along these lines. Neither Weak nor Strong Epistemic Teleology can be saved by switching one's account of what the intrinsic epistemic values or ultimate epistemic goals are in any plausible way.

5.   Some Loose Ends and the Way Forward

So we have strong reasons to reject Strong Epistemic Teleology. Nevertheless, my task of fully defending a clear alternative is hardly complete. For there are some loose ends left over by my arguments that are worth mentioning. These will set some of the goals for the next chapters.

One thing that needs to be done is to articulate more completely the core ideas behind epistemic Kantianism as an account of the fundamental epistemic norm. In this chapter, I have taken it for granted that we all understand what it is to respect the truth with respect to a proposition. Indeed, I also assumed a particular view about what respect for the truth is in giving my arguments. Unsurprisingly, then, one might have at least two lingering complaints.

The first complaint is about the contentfulness of the requirement of respect for truth in belief with respect to the target proposition. To see the worry, consider a closely related problem about Kantian ethics. While Parfit (2011) is sympathetic to the idea that persons essentially call for respect, and that this is the fundamentally correct response to their value, he was also skeptical that one could derive any interesting normative conclusions from this idea. He wrote:

> The requirement to respect all persons is one of Kant's greatest contributions to our moral thinking. But it does not tell us how we ought to act. […] Some writers suggest that
>
> (B) it is wrong to treat people in ways that are incompatible with respect for them.
>
> Some wrong acts are clearly incompatible with respect for persons. Kant's examples are: disgraceful or humiliating punishments, ridicule, defamation, and acts that display arrogance or contempt. But Kant's formula is intended to cover all wrong acts, and most wrong acts do not treat people in such disrespectful ways.
> All wrong acts, some writers suggest, are in a wider sense incompatible with respect for persons. On this suggestion, (B) would not be a useful claim. As before, to decide whether some act would be in this wider sense incompatible with respect for persons, we would first have to decide whether this act would be wrong. If this act would *not* be wrong, it would be compatible with respect for persons. As both Kant and Sidgwick warn, moral philosophers often make claims that seem to give us 'valuable information' but really tell us only that acts are wrong if they are wrong.[204]

One might have the same concern about my epistemic Kantian view. One might think that all we can usefully say about what it is to respect the truth in some case is to form the doxastic attitude that is best supported by the epistemic reasons. And that would not be helpful.

Now, I assumed a particular account of what it is to respect the truth earlier that avoids this objection. I said that to respect the truth with respect to P is to successfully fully constrain one's

---

theoretical deliberation with respect to the question of whether P by the injunction expressed by the Conditional T-Goal. This is a substantive claim. As I've suggested, the conjunction of it and the view that the requirement to respect the truth is the most fundamental epistemic norm implies the falsity of some widely accepted theories in epistemology. So it severely constrains what theories we can accept. But this simply brings us to a second complaint—viz., that I have not done anything to support this particular account of what it is to respect the truth.

In reply, I can say this: Chapters 5 and 6 will be devoted to addressing these complaints systematically. By the end of these chapters, we will have a complete understanding both of the epistemic Kantian view insofar as it is a view about the fundamental epistemic norm, as well as insight into how this view solves the problems of deontic significance for epistemic rationality.

A second loose end is left dangling from my attack on Weak Epistemic Teleology. What I mainly defended in this chapter was the conclusion that the fundamentally correct response to the epistemic value of truth is non-teleological and, more specifically, the respect-involving one as suggested by epistemic Kantianism. But this hardly constituted a defense of the claim that *all* important facts about epistemic value can be explained in the epistemic Kantian's way. One might, then, have the lingering worry that even if the epistemic value of truth can be understood as the epistemic Kantian wants, there are other epistemic values that cannot be so understood.

This worry will leave one wondering whether the requirement to respect the truth really could be the *only* fundamental epistemic norm, even if it is *a* fundamental epistemic norm. After all, if there are other epistemic values to which the fundamentally correct response is *not* the Kantian one of respect, and their epistemic value cannot be derived in some way from the fact that the intrinsic value of truth *can* be so understood, one might think the fundamentally correct responses to these other epistemic values will give rise to further deontic constraints on our doxastic attitudes and theoretical reasoning—specifically, by the Value-Driven Constraint on the Criteria of Quality for Reasons exploited in §4.2. If so, that would call into question the idea of using the epistemic Kantian account of the epistemic value of truth to provide a full underpinning of the deontic significance of epistemic rationality. If other epistemic values give rise to different constraints, they may conflict with the respect-based one, and thereby possibly reintroduce the problems of deontic significance.

Chapter 7 will be devoted to addressing this concern. Together with Chapters 5 and 6, it will yield a complete defense of epistemic Kantianism as an account of all deontic and evaluative facts in epistemology. The implications of this defense for epistemology as a whole will then be explored in Chapter 8. With this bigger picture in mind, I'll turn to tie up the first loose end.

REFERENCES

Alston, W.　1988.　"An Internalist Externalism." *Synthese* 74.3: 265–83.

_____.　2002a. "Plantinga, Naturalism, and Defeat" in Beilby, J. (ed.) <u>Naturalism Defeated?</u>. Ithaca: Cornell University Press.

_____.　2002b.　"Sellars and the 'Myth of the Given'." *Philosophy and Phenomenological Research* 65: 69–86.

_____.　2005.　<u>Beyond "Justification"</u>. Ithaca: Cornell University Press.

Anderson, E.　1993.　<u>Value in Ethics and Economics</u>.　Cambridge, MA: Harvard University Press.

Armstrong, D.　1973.　<u>Belief, Truth and Knowledge</u>.　Cambridge: Cambridge University Press.

_____.　1997.　<u>A World of States of Affairs</u>.　Cambridge: Cambridge University Press.

Austin, J. L.　1970.　<u>Philosophical Papers</u>.　Oxford: Clarendon Press.

Bach, K.　1985.　"A Rationale for Reliabilism." *The Monist* 68: 246–63.

Baehr, J.　2009.　"Is There a Value Problem?" in (eds.) Haddock, A., Millar, A., and Pritchard, D. (eds.) <u>Epistemic Value</u>. Oxford: Oxford University Press.

Bayne, T.　2009.　"Perception and the Reach of Phenomenal Content." *Philosophical Quarterly* 59: 385–404.

Bedke, M.　2011.　"Passing the Deontic Buck" in *Oxford Studies in Metaethics* 6.　Oxford Oxford University Press.

Bengson, J.　Ms.　<u>The Intellectual Given</u>.　Ph.D. Dissertation, University of Texas, Austin.

Berker, S.　**Ms**$_1$.　"Epistemic Teleology and the Separateness of Propositions."

_____.　Ms$_2$.　"The Rejection of Epistemic Consequentialism."

Bergmann, M.　2006.　<u>Justification without Awareness</u>.　Oxford: Oxford University Press.

Bird, A.　2007.　"Justified Judging." *Philosophy and Phenomenological Research* 74.1: 81–110.

Boghossian, P.　2008.　<u>Content and Justification</u>.　Oxford: Oxford University Press.

_____.　2009.　"Virtuous Intuitions: Comments on Lecture 3 of Ernest Sosa's *A Virtue Epistemology*." *Philosophical Studies* 144: 111–119.

BonJour, L.　1980.　"Externalist Theories of Empirical Knowledge." *Midwest Studies in Philosophy* 5: 53–73.

_____.   1985.   The Structure of Empirical Knowledge.   Cambridge MA: Harvard University Press.

BonJour, L. and Sosa, E.   2003.   Epistemic Justification. Oxford: Blackwell.

Brandt, R.   1979.   A Theory of the Good and the Right. Oxford: Oxford University Press.

Bratman, M.   1992.   "Acceptance and Practical Reasoning in a Context." *Mind* 101: 327–41.

Brook, R.   1991.   "Agency and Morality." *Journal of Philosophy* 88: 190–212.

Broome, J.   1999.   "Normative Requirements." *Ratio* 12: 398–419.

_____.   2007a. "Requirements" in Hommage a Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz. http://www.fil.lu.se/hommageawlodek/

_____.   2007b. "Wide or Narrow Scope?" *Mind* 116: 359–370.

_____.   2008.   "Is Rationality Normative?" *Disputatio* 11: 153–71.

_____.   Ms.   Rationality through Reasoning.

Brueckner, A. 2005.   "Knowledge, Evidence and Skepticism according to Williamson." *Philosophy and Phenomenological Research* 70: 436-43.

_____.   2009. "E=K and Perceptual Knowledge" in Greenough, Patrick and   Pritchard, Duncan (eds.) Williamson on Knowledge. Oxford: Oxford University Press.

Chang, R. (ed.) 1997.   Incommensurability, Incomparability, and Practical Reason. Cambridge, MA: Harvard University Press.

Chisholm, R.   1957.   Perceiving: A Philosophical Study. Ithaca: Cornell University Press.

Christensen, D. 2010.   "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81: 185–215.

Chudnoff, E.   2011.   "What Intuitions Are Like." *Philosophy and Phenomenological Research* 72: 625–654.

Cohen, L. J.   1992.   An Essay on Belief and Acceptance. Oxford: Oxford University Pres.

Cohen, S.   1984.   "Justification and Truth." *Philosophical Studies* 46.3: 279–95.

Comesaña, J.   2010.   "Evidentialist Reliabilism." *Nous* 44.4: 571-600.

Conee, E. and Feldman, R.   2011. "Replies" in Dougherty, T. (ed.) Evidentialism and its Discontents. Oxford: Oxford University Press.

Cullison, A.   2009.   "Moral Perception." *European Journal of Philosophy* 18.2: 159–75.

Dancy, J.   1992.   "Externalism for Internalists." *Philosophical Issues* 2: 93–114.

_____.   2000.   Practical Reality. Oxford: Oxford University Press.

_____.   2004a.   Ethics without Principles. Oxford: Oxford University Press.

_____.   2004b. "Two Ways of Explaining Actions" in Hyman, J. and Steward, H. (eds.) Agency and Action. Cambridge: Cambridge University Press.

Darwall, S.   1977.   "Two Kinds of Respect." *Ethics* 88: 36–49.

_____.   1983.   Impartial Reason. Ithaca: Cornell University Press.

_____.   2006.   The Second-Person Standpoint.   Cambridge MA: Harvard University Press.

Davidson, D.   1963.   "Actions, Reasons, and Causes." *Journal of Philosophy* 60: 685–700.

_____.   1982.   "Paradoxes of Irrationality" in Wollheim, R. and Hopkins, J. (eds.) Philosophical Essays on Freud. Cambridge: Cambridge University Press.

_____.   1984.   "Incoherence and Irrationality." *Dialectica* 39: 345–54.

_____.   2004.   Problems of Rationality. Oxford: Oxford University Press.

De Sousa, R.   1990.   The Rationality of Emotion. Cambridge MA: The MIT Press.

Dretske, F.   1969.   Seeing and Knowing. Chicago: University of Chicago Press.

_____.   1971.   "Conclusive Reasons." *Australasian Journal of Philosophy* 49: 1–22.

_____.   1981.   Knowledge and the Flow of Information.   Cambridge MA: The MIT Press.

_____.   1991.   "Two Conceptions of Knowledge." *Grazer Philosophiche Studien*: 15–30.

_____.   2000.   "Perception, Knowledge and Belief." Cambridge: Cambridge University Press.

Dummett, M. 1991.   The Logical Basis of Metaphysics. Cambridge MA: Harvard University

Press.

Dunlosky, J. and Bjork, R. A.   2008.   Handbook of Metamemory and Memory.   New York: Psychology Press (Taylor and Francis).

Efklides, A. and Misailidi, P.   2010.   Trends and Prospects in Metacognition Research.   New York: Springer.

Engel, M.   1992.   "Personal and Doxastic Justification in Epistemology."   *Philosophical Studies* 67: 133–50.

Fales, E.   1996.   A Defense of the Given.   Lanham, MD: Rowman and Littlefield.

Fantl, J. and McGrath, M.   2002.   "Evidence, Pragmatics and Justification."   *Philosophical Review* 111: 67–94.

————.   2009.   Knowledge in an Uncertain World.   Oxford: Oxford University Press.

Feldman, R.   2005.   "Justification Is Internal" in Steup, M. and Sosa, E. (eds.) Contemporary Debates in Epistemology.   Oxford: Blackwell.

Firth, R.   1981.   "Epistemic Merit, Intrinsic and Instrumental."   *Proceedings and Addresses of the American Philosophical Association* 55: 5–23.

Fodor, J.   1987.   Psychosemantics.   Cambridge MA: The MIT Press.

————.   1990.   A Theory of Content and Other Essays.   Cambridge MA: The MIT Press.

————.   2004a.   "Having Concepts: A Brief Refutation of the Twentieth Century."   *Mind & Language* 19: 29–47.

————.   2004b.   "Reply to Commentators."   *Mind & Language* 19: 99–112.

————.   2008.   LOT 2: The Language of Thought Revisited.   Oxford: Oxford University Press.

Foley, R.   1987.   The Theory of Epistemic Rationality.   Cambridge MA: Harvard University Press.

————.   1993.   Working without a Net.   Oxford: Oxford University Press.

————.   2004.   "A Trial Separation between the Theory of Knowledge and the Theory of Justified Belief." in Greco, J. (ed.) Ernest Sosa and His Critics.   Oxford: Blackwell.

Foot, P.   1978.   Virtues and Vices.   Berkeley: University of California Press.

————.   1985.   "Utilitarianism and the Virtues."   *Mind* 94: 196–209.

————.   2003.   Moral Dilemmas.   Oxford: Oxford University Press.

Frankfurt, H.   1971.   "Freedom of the Will and the Concept of a Person."   *Journal of Philosophy* 68: 5–20.

Fumerton, R.   1980.   "Induction and Reasoning to the Best Explanation."   *Philosophy of Science* 47: 589–600.

————.   1990.   Reason and Morality.   Ithaca: Cornell University Press.

————.   1995.   Metaepistemology and Skepticism.   Lanham MD: Rowman and Littlefield.

————.   2001.   "Epistemic Justification and Normativity" in Steup, M. (ed.) Knowledge, Truth and Duty.   Oxford: Oxford University Press.

————.   2007.   "Review of Justification without Awareness."   *Notre Dame Philosphical Reviews* http://ndpr.nd.edu/review.cfm?id=9104.

Gabbay, M.   Ms.   "We Can Be in Harmony with Classical Logic."

Geach, P. T.   1956.   "Good and Evil."   *Analysis* 17: 33–42.

Gibbard, A.   1990.   Wise Choices, Apt Feelings.   Cambridge MA: Harvard University Press.

Ginet, C.   1975.   Knowledge, Perception and Memory.   Dordrecht: D. Reidel.

Goldman, A.   1967.   "A Causal Theory of Knowing." *Journal of Philosophy* 73: 771–91.

————.   1979.   "What Is Justified Belief?" in Pappas, G. (ed.) *Justification and Knowledge*.   Dordrecht: Reidel.

————.   1980.   "The Internalist Conception of Justification."   *Midwest Studies in Philosophy* 5: 27–52

————.   1986.   Epistemology and Cognition.   Cambridge MA: Harvard University Press.

_____. 1988. "Strong and Weak Justification." *Philosophical Perspectives* 2: 51–69.

_____. 1999a. "Internalism Exposed." *Journal of Philosophy* 96: 271–93.

_____. 1999b Knowledge in a Social World. Oxford: Oxford University Press.

_____. 2002. "The Unity of the Epistemic Virtues" in Pathways to Knowledge. Oxford: Oxford University Press.

_____. 2009. "Internalism, Externalism and the Architecture of Justification." *Journal of Philosophy* 106(6): 1–30.

_____. 2011. "Toward a Synthesis of Reliabilism and Evidentialism?" in Dougherty, T. (ed.) Evidentialism and its Discontents. Oxford: Oxford University Press.

Goldman, A. and Olsson, E. 2009. "Reliabilism and the Value of Knowledge" in (eds.) Haddock, A., Millar, A., and Pritchard, D. (eds.) Epistemic Value. Oxford: Oxford University Press.

Greco, J. 2005. "Justification Is Not Internal" in Steup, M. and Sosa, E. (eds.) Contemporary Debates in Epistemology. Oxford: Blackwell.

_____. 2009. "The Value Problem" in (eds.) Haddock, A., Millar, A., and Pritchard, D. (eds.) Epistemic Value. Oxford: Oxford University Press.

_____. 2010. Achieving Knowledge. Cambridge: Cambridge University Press.

Grimm, S. 2009. "Epistemic Normativity" in Haddock, A., Millar, A., and Pritchard, D. (eds.) Epistemic Value. Oxford: Oxford University Press.

Hawthorne, J. 2004. Knowledge and Lotteries. Oxford: Oxford University Press.

Hawthorne, J. and Stanley, J. 2008. "Knowledge and Action." *Journal of Philosophy* 105: 571–90.

Hieronymi, P. 2005. "The Wrong Kind of Reason." *Journal of Philosophy* 102: 437–57.

Howard-Snyder, F. 1997. "The Rejection of Objective Consequentialism." *Utilitas* 9: 241–48.

_____. 2005. "It's the Thought that Counts." *Utilitas* 17: 265–81.

Heuer, U. 2004. "Raz on Values and Reasons" in Wallace, R. J., Pettit, P., Scheffler, S., and Smith, M. (eds.) Reason and Value. Oxford: Oxford University Press.

Huemer, M. 2001. Skepticism and the Veil of Perception. Lanham MD: Rowman and Littlefield.

_____. 2007. "Compassionate Phenomenal Conservatism." *Philosophy and Phenomenological Research* 74: 30–55.

Hurka, T. 2001. Virtue, Vice and Value. Oxford: Oxford University Press.

Hursthouse, R. 1991. "Arational Actions." *Journal of Philosophy* 88: 57–68.

Jackson, A. 2011. "Appearances, Rationality and Justified Belief." *Philosophy and Phenomenological Research* 82.3: 564–93.

Jackson, F. 1991. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101: 461–82.

Jacobson, D. 2005. "Seeing by Feeling: Virtues, Skills, and Moral Perception." *Ethical Theory and Moral Practice* 8: 387–409.

Johnston, M. 2011. "On a Neglected Epistemic Virtue." *Philosophical Issues* 21: 165–218.

Kagan, S. 1989. The Limits of Morality. Oxford: Clarendon Press.

_____. 1998. "Rethinking Intrinsic Value." *The Journal of Ethics* 2: 277–97.

Kelly, T. 2002. "The Rationality of Belief and Some Other Propositional Attitudes." *Philosophical Studies* 110: 163–96.

_____. 2003. "Epistemic Rationality as Instrumental Rationality: A Critique." *Philosophy and Phenomenological Research* 66: 612–40.

_____. 2007. "Evidence and Normativity." *Philosophy and Phenomenological Research* 75: 465–74.

Kelp, C. Ms. "Knowledge: The Safe-Apt View."

Kim, J. 1976. "Events as Property Exemplifications" in Brand, M. and Walton, D. (eds.) Action Theory. Dordrecht: Reidel.

Koksvik, O.   Ms.   "Intuition, Belief, and Rational Criticizability."

Kolodny, N.   2005.   "Why Be Rational?" *Mind* 114: 509–63.

_____.   2008a.   "How Does Coherence Matter?" *Proceedings of the Aristotelian Society* 107:1: 229–263.

_____.   2008b.   "The Myth of Practical Consistency." *European Journal of Philosophy* 16:3: 366–402.

_____.   2008c.   "Why Be Disposed to Be Coherent?" *Ethics* 118:3: 437–63.

Kornblith, H.   2008.   "Knowledge Needs No Justification" in Smith, Q. (ed.) Epistemology: New Essays. Oxford: Oxford University Press.

Korsgaard, C.   1983.   "Two Distinctions in Goodness." *The Philosophical Review* 92.2: 169–95.

_____.   1986.   "Skepticism about Practical Reason." *Journal of Philosophy* 83: 5–25.

Kvanvig, J.   2003.   The Value of Knowledge and the Pursuit of Understanding. Cambridge: Cambridge University Press.

Langton, R.   2007.   "Objective and Unconditioned Value."

Laurence, S. and Margolis, E. 1999. Concepts: Core Readings. Cambridge MA: The MIT Press.

Lehrer, K.   1974.   Knowledge. Oxford: Clarendon Press.

Leite, A.   2007.   "Epistemic Instrumentalism and Reasons for Belief." *Philosophy and Phenomenological Research* 75: 456–64.

Lewis, D. K.   1983.   "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343–77.

Littlejohn, C.   2009.   "The Externalist's Demon." *Canadian Journal of Philosophy* 39: 399–434.

_____.   2011.   "Reasons and Belief's Justification" in Reisner, A. and Steglich-Petersen (eds.) Reasons for Belief. Cambridge: Cambridge University Press.

Lord, E.   2008.   "Dancy on Acting for the Right Reason." *Journal of Ethics and Social Philosophy.* http://www.jesp.org/articles/download/ErrolLordDiscussion.pdf

_____.   2010.   "Having Reasons and the Factoring Account." *Philosophical Studies* 149: 283–96.

_____.   Ms1.   "Correctly Responding to Reasons and the Normativity of Rationality." http://www.princeton.edu/~edlord/Site/papers_files/correctly%20responding%20to%20reasons.pdf

_____.   Ms2.   "Explaining Wide-Scope Requirements." http://www.princeton.edu/~edlord/Site/papers_files/explaining%20wide-scope%20requirements.pdf

Lynch, M.   2004.   True to Life: Why Truth Matters. Cambridge MA: MIT Press.

Lyons, J.   2005.   "Perceptual Belief and Nonexperiential Looks." *Philosophical Perspectives* 19: 237–56.

Margolis, E.   1998.   "How to Acquire a Concept." *Mind and Language* 13: 347–69.

Markie, P.   2008.   "Review of Justification without Awareness." *The Philosophical Quarterly* 58: 550–55.

Markovits, J.   2011.   "Why Be an Internalist about Reasons?" *Oxford Studies in Metaethics* 6. Oxford: Oxford University Press.

McDowell, J.   1998.   Value, Mind and Reality. Cambridge MA: Harvard University Press.

_____.   2011.   Perception as a Capacity for Knowledge. Milwaukee: Marquette University Press.

McGee, V.   1985.   "A Counterexample to *Modus Ponens*." *Journal of Philosophy* 82: 462–71.

McGrath, S.   2004.   "Moral Knowledge by Perception." *Philosophical Perspectives* 18: 209–28.

McHugh, C.   Forthcoming. "Exercising Doxastic Freedom." *Philosophy and Phenomenological Research.*

McKeever, S. and Ridge, M. 2006. Principled Ethics. Oxford: Oxford University Press.

Millar, A.   1991.   Reasons and Experience. Oxford: Oxford University Press.

_____.   2004.   Understanding People. Oxford: Oxford University Press.

Moore, E.   2007.   "Objective Consequentialism, Right Actions, and Good People."

Philosophical Studies 133: 83–94.

Nagel, J.   2010.   "Epistemic Anxiety and Adaptive Invariantism." *Philosophical Perspectives* 24: 407–435.

_____.   Ms.   "Factual Memory, Internalism, and Metacognition." Available *via* this URL: http://www.utm.utoronto.ca/~nageljen/trivia.pdf

Nagel, T.   1970.   The Possibility of Altruism. Princeton: Princeton University Press.

Nozick, R.   1981.   Philosophical Explanations. Cambridge MA: Belknap Press.

Otsuka, M.   2009.   "The Kantian Argument for Consequentialism." *Ratio* 22: 41–58.

Owens, D.   2000.   Reason without Freedom. London: Routledge.

Parfit, D.   2001.   "Rationality and Reasons" in Dan Egonsson, Björn Petterson & Toni Rønnow-Rasmussen (eds.) Exploring Practical Philosophy. Aldershot: Ashgate.

_____.   2011.   On What Matters. Oxford: Oxford University Press.

Piller, C.   2006.   "Content-Related and Attitude-Related Reasons for Preferences." *Philosophy* 81: 155–81.

Plantinga, A.   1993.   Warrant and Proper Function. Oxford: Oxford University Press.

Pritchard, D.   2005.   Epistemic Luck. Oxford: Oxford University Press.

_____, Haddock, A. and Millar, A. 2011.   The Nature and Value of Knowledge: Three Investigations. Oxford: Oxford University Press.

Pryor, J.   2000.   "The Skeptic and the Dogmatist." *Nous* 34: 517–49.

_____.   2001.   "Highlights of Recent Epistemology." *British Journal for the Philosophy of Science* 52: 95–124.

_____.   2005.   "What's Wrong with Moore's Argument?" *Philosophical Issues* 14: 349–378.

Rabinowicz, W. and Rønnow-Rasmussen, T. 1999. "A Distinction in Value: Intrinsic and For its Own Sake." *Proceedings of the Aristotelian Society* 100: 33–49.

Raz, J.   2002.   Engaging Reason. Oxford: Oxford University Press.

_____.   2005.   "The Myth of Instrumental Rationality." *Journal of Ethics and Social Philosophy* 1.1: 1–28.

Rawls, J.   1971.   A Theory of Justice. Oxford: Oxford University Press.

Read, S.   2000.   "Harmony and Autonomy in Classical Logic." *Journal of Philosophical Logic* 29: 123–54.

                2010.   "General-Elimination Harmony and the Meaning of the Logical Constants." *Journal of Philosophical Logic* 39: 557–76.

Reisner, A.   2011a. "Is There Reason to Be Theoretically Rational?" in Reisner, A. and Steglich-Petersen (eds.) Reasons for Belief. Cambridge: Cambridge University Press.

_____.   2011b. "Unifying the Requirements of Rationality." *Philosophical Explorations* 12: 243–260.

Ridge, M.   Forthcoming. "Getting Lost on the Road to Larissa." *Nous*.

Rosen, G.   2004.   "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18: 295–313.

Ross, J.   2009.   "How to Be a Cognitivist about Practical Reason." *Oxford Studies in Metaethics* 4. Oxford: Oxford University Press.

_____.   Forthcoming. "Rationality, Normativity and Commitment." *Oxford Studies in Metaethics* 7. Oxford: Oxford University Press.

Scanlon, T. M. 1998.   What We Owe to Each Other. Cambridge MA: Harvard University Press.

_____.   2008.   Moral Dimensions. Cambridge MA: Harvard University Press.

_____.   2011.   "The Unity of the Normative." *Philosophical Studies* 54: 443–50.

Schaffer, J.   2003.   "Overdetermining Causes." *Philosophical Studies* 114: 23–45.

Schechter, J.   2011.   "Rational Self-Doubt and the Failure of Closure." *Philosophical Studies*

Schechter, J. and Enoch, D. 2006. "Meaning and Justification: the Case of Modus Ponens." *Nous* 40: 687–715.

Schnee, I.   Ms$_1$.   "Reasoning from the Facts."

_____.　　Ms₂.　"Duplicate Reasons."

Schroeder, M.　2004.　"The Scope of Instrumental Reason." *Philosophical Perspectives* 18: 337–364.

_____.　　2005.　"Realism and Reduction: The Quest for Robustness." *Philosophers' Imprint* 5.1. <www.philosophersimprint.org/005001/>

_____.　　2007.　Slaves of the Passions. Oxford: Oxford University Press.

_____.　　2009.　"Means-End Coherence, Stringency and Subjective Reasons." *Philosophical Studies* 143: 223–248.

_____.　　2010.　"Value and the Right Kind of Reason." *Oxford Studies in Metaethics* 5. Oxford: Oxford University Press.

_____.　　2011.　"What Does It Take to 'Have' a Reason?" in Reisner, A. and Steglich-Petersen, A. (eds.) Reasons for Belief. Cambridge: Cambridge University Press.

_____.　　Forthcoming.　"States, Withholding and Pragmatic Encroachment on Knowledge." *Philosophical Studies*.

_____.　　Ms₁.　"The Ubiquity of State-Given Reasons."

_____.　　Ms₂.　"What Makes Reasons Sufficient?"

Setiya, K.　　2007.　"Cognitivism about Instrumental Rationality." *Ethics* 117: 649–73.

Shah, N.　　2003.　"How Truth Governs Belief." *Philosophical Review* 112: 447–482.

_____ and Velleman, J. D.　2005.　"Doxastic Deliberation." *Philosophical Review* 114: 497–534.

_____.　　2006.　"A New Argument for Evidentialism." *Philosophical Quarterly* 56: 481–98.

Sher, G.　　2009.　Who Knew?. Oxford: Oxford University Press.

Sider, T.　　2009.　"Ontological Realism" in Chalmers, D., Manley, D. and Wasserman, R. (eds.) Metametaphysics. Oxford: Oxford University Press.

Skorupski, J.　2007.　"Buck-Passing about Goodness" in Rønnow-Rasmussen, Petersson, Jesefsson, and Egonsson (eds.) Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz. www.fil.lu.se/hommageawlodek .

_____.　　2011.　The Domain of Reasons. Oxford: Oxford University Press.

Slote, M.　　1995.　"Agent-Based Virtue Ethics." *Midwest Studies in Philosophy* 20: 83–101.

Smith, H.　　1976.　"Dated Rightness and Moral Imperfection." *Philosophical Review* 85: 449–87.

_____.　　1983.　"Culpable Ignorance." *Philosophical Review* 92: 543–71.

_____. Forthcoming.　"The Moral Clout of Reasonable Beliefs." *Oxford Studies in Normative Ethics* 2. Oxford: Oxford University Press.

Smith, M.　　1994.　The Moral Problem. Oxford: Blackwell.

_____.　　1995.　"Internal Reasons." *Philosophy and Phenomenological Research* 55: 109–131.

_____.　　2007.　"Is There a Nexus between Reasons and Rationality?" *Poznan Studies in the Philosophy of the Sciences and Humanities* 94: 279–298.

Sosa, E.　　1991.　Knowledge in Perspective. Oxford: Oxford University Press.

_____.　　2001.　"For the Love of Truth?" in Fairweather, A. and Zagzebski, L. (eds.) Virtue Epistemology: Essays on Epistemic Virtue and Responsibility. Oxford: Oxford University Press.

_____.　　2003.　"The Place of Truth in Epistemology" in DePaul, M. and Zagzebski, L. (eds.) Intellectual Virtue. Oxford: Oxford University Press.

_____.　　2007.　A Virtue Epistemology: Apt Belief and Reflective Knowledge. Oxford: Oxford University Press.

_____.　　2009.　"Replies to Commentators on *A Virtue Epistemology*." *Philosophical Studies* 144: 137–47.

_____.　　2010.　Knowing Full Well. Princeton: Princeton University Press.

_____.　　Ms.　"Synchronism."

Sylvan, K.　　Ms₁.　"How Practical Rationality Is Not Interestingly Belief-Relative."

_____.　　Ms₂.　"Inferentialism, Pragmatism, and Circularity."

_____.　　Ms₃.　"The Role of Fittingness in the Analysis of Value."

_____.   Ms₄.   "The Wide-Scoper's Bootstrapping Problem?"

Swank, C.   2000.   "Epistemic Vice" in Axtell, G. (ed.) <u>Knowledge, Belief, and Character</u>. Lanham: Rowman & Littlefield.

Tannenbaum, J.2008.   "Categorizing Goods." *Oxford Studies in Metaethics* 5. Oxford: Oxford University Press.

Thomson, J.   1986.   <u>Rights, Restitution and Risk</u>. Cambridge MA: Harvard University Press.

                1992.   <u>The Realm of Rights</u>. Cambridge MA: Harvard University Press.

_____.   1997.   "The Right and the Good." *Journal of Philosophy* 94: 273–298.

_____.   2008.   <u>Normativity</u>. Chicago: Open Court.

Turri, J.   2009.   "The Ontology of Epistemic Reasons." *Nous* 43.3: 490–512.

_____.   2010.   "On the Relationship between Propositional and Doxastic Justification." *Philosophy and Phenomenological Research* 80.2: 312–26.

_____.   2011.   "Believing for a Reason." *Erkenntnis* 74: 383–397.

Unger, P.   1975.   <u>Ignorance</u>. Oxford: Oxford University Press.

Von Wright, G. 1963.   <u>The Varieties of Goodness</u>. New York: The Humanities Press.

Way, J.   2010.   "Two Accounts of the Normativity of Rationality." *Journal of Ethics and Social Philosophy.*

_____.   2011.   "The Symmetry of Rational Requirements." *Philosophical Studies* 155: 227–239.

Weatherson, B. 2008.   "Deontology and Descartes's Demon." *Journal of Philosophy* 105: 540–69.

Wedgwood, R.   2006.   "The Normative Force of Reasoning." *Nous* 40.4: 660–86.

_____.   Ms.   "The Right Thing to Believe."

Whitcomb, D.   Ms.   <u>An Epistemic Value Theory</u>. Ph.D. Dissertation, Rutgers University.

Williams, B.   1981.   "Internal and External Reasons" in <u>Moral Luck</u>. Cambridge: Cambridge University Press.

_____.   1985.   <u>Ethics and the Limits of Philosophy</u>. London: Fonatana.

Williamson, T. 2000.   <u>Knowledge and its Limits</u>. Oxford: Oxford University Press.

_____.   2003.   "Blind Reasoning." *Proceedings of the Aristotelian Society*, *Supplementary Volume* 77: 249–93.

_____.   2005.   "Replies to Commentators" *Philosophy and Phenomenological Research* 70: 468–91.

_____.   2007.   <u>The Philosophy of Philosophy</u>. Oxford: Blackwell.

_____.   2009.   "Replies" in Greenough, P. and Pritchard, D. (eds.) <u>Williamson on Knowledge</u>. Oxford: Oxford University Press.

Wright, C.   2004.   "Warrant for Nothing? (and Foundations for Free)?" *Aristotelian Society Supplement* 78: 167–245.

Zagzebski, L.   1996.   <u>Virtues of the Mind</u>. Cambridge: Cambridge University Press.

Zimmerman, M. 1997."A Plea for Accuses." *American Philosophical Quarterly* 34.2: 229–43.

_____.   2001.   <u>The Nature of Intrinsic Value</u>. Lanham: Rowman & Littlefield.

_____.   2002.   "Taking Luck Seriously." *Journal of Philosophy* 99: 553–73.

_____.   2008.   <u>Living with Uncertainty: The Moral Significance of Ignorance</u>. Cambridge: Cambridge University Press.