

Internalism Re-explained¹

Ralph Wedgwood

1. An intuitive argument for internalism

Consider two possible worlds, w_1 and w_2 . In both worlds, you have exactly the same experiences, apparent memories, and intuitions, and in both worlds you go through exactly the same processes of reasoning, forming, maintaining, and revising exactly the same beliefs in exactly the same ways. It seems clear that these two worlds are also exactly alike with respect to which of your beliefs are rational and which are irrational. Now suppose that in w_1 you are bedevilled by an evil demon who ensures that many of your experiences are misleading, with the result that many of the beliefs that you hold in w_1 are false. In w_2 , on the other hand, almost all your experiences are veridical, with the result that almost all the beliefs that you hold in w_2 are true. Intuitively, this makes no difference. Exactly the same beliefs are rational in both worlds.

This intuition seems to support an “internalist” conception of rational belief.² According to this conception, the rationality of a belief supervenes purely on “internal facts” about what is present in the thinker’s mind – in this example, on facts that hold in *both* these two possible worlds w_1 and w_2 , not on facts about the external world that vary between w_1 and w_2 . That is, there cannot be any difference between the worlds w_1 and w_2 with respect to which of the thinker’s beliefs are rational in those worlds unless there is also a difference with respect to the “internal facts” about what is present in the thinker’s mind in those worlds.

Since this intuition seems to support such an “internalist” conception of rational belief, it poses a *prima facie* problem for the rival “externalist” conception.³ This problem for externalism has come to be known as the “new evil demon problem”. Externalists have attempted various different ways of solving this problem.⁴ I shall not examine these attempted solutions here. Instead, the main goal of this essay is to offer a systematic defence of the view that (contrary to what externalists have claimed) this intuitive argument for internalism is in fact perfectly sound.

¹ Earlier versions of this paper were presented at talks at the University of Otago, the Center for the Study of Mind in Nature (CSMN) at the University of Oslo, the University of Stockholm, and Rutgers University. I am grateful to the members of those audiences for helpful comments.

² For an early statement of this kind of argument, see Cohen (1984). I invoked this argument at the beginning of an earlier essay of mine (Wedgwood 2002); the goal of the present chapter is to restate the basic position of that earlier essay in a new and improved form.

³ For some leading exponents of such an externalist position, see e.g. Williamson (2000, Chap. 9), and Goldman (1979).

⁴ For example, see the response to the new evil demon problem of Goldman (1986, 107–09). A number of externalists – such as Sosa (1994) and Comesaña (2002) – effectively concede that the letter of this argument is correct, while disputing its spirit – in particular, by suggesting that our actual use of term ‘rational’ refers to those processes and methods of reasoning that are *actually* reliable. On this view, the term ‘rational’ as used by the thinkers of demon-worlds would refer to quite different processes of reasoning from those that the term refers to when used by the thinkers of normal worlds. In this chapter, I shall assume that this view is incorrect, and that the reference of the term ‘rational’ does not vary between worlds in this way.

Indeed, intuitively, internalism seems to articulate a completely *general* feature of rationality. It is not just rational *belief-states* that have this feature: the same feature seems to hold of rational *processes of belief revision*; and it also seems to hold of rational mental events and mental states of other kinds, such as rational decisions and intentions. Whenever we assess any process of reasoning or mental state or event as rational or irrational, we are assessing it on the basis of its relation to the mental events and states that are present in thinker's mind – not on the basis of its relation to facts about the external world that could vary while those mental events and states remained unchanged.

In this way, internalism articulates a completely general feature of rationality. However, in the discussion that follows, I shall mostly focus on rational belief: I shall only occasionally touch on rational processes of belief revision, and I shall completely ignore the phenomena of rational decisions and intentions. In my view, all the arguments that I shall offer here can be generalized so that they apply to rational processes of reasoning, and to rational decision and intention, as well as to rational belief; but I shall not take the time to justify this view here.⁵

More importantly, internalism seems not only to articulate a general feature of rationality; it also seems to articulate a *special* feature of rationality, in contrast to other ways of evaluating beliefs and the like. All the other terms that can be used for normatively evaluating beliefs – for example, as “correct” or “incorrect”, “right” or “wrong”, and so on – are externalist evaluations. What is distinctive of “rationality” (at least as the term is most commonly used by philosophers) is that it is an internalist evaluation.

It seems plausible that the concept of what is *rationally required* of a particular thinker at a particular time is in effect a kind of ‘ought’. So there should be at least one kind of ‘ought’ that also supervenes on internal facts about the mental events and states that are present in the relevant thinker's mind at the relevant time. As a matter of fact, it does seem that there is such a way of using ‘ought’. Suppose that (given all my prior background beliefs, and the experiences and memories that I am currently having, and so on) I ought to believe that I am currently sitting on a bus travelling from London to Oxford. Now suppose that there is another thinker who has exactly the same experiences, memories, and prior background beliefs that I have, but is being deceived by an evil demon, so that in fact almost all of his sensory experiences and apparent memories are misleading illusions, and not reliable guides to how things are in his environment. Nonetheless, it seems to me that there is a way of using the word ‘ought’ in which it is true to say that since his beliefs, experiences, and memories are just the same as mine, he also ought to believe that he is sitting on a bus from London to Oxford. This certainly seems the most rational thing for him to believe. So how could it fail to be what he in some sense ought to believe?

In this way, internalism about rationality seems to be supported by an intuitively compelling argument. As we shall see, however, several questions can be raised about this argument.

⁵ A question could be raised about rational *actions*: suppose that in world w_1 , just as you are about to execute a rational intention to tie your shoe laces, the evil demon intervenes, simultaneously giving you an experience as of your tying your shoe laces, while paralyzing your body (so that you do not execute this intention at all) – while in world w_2 , you rationally tie your shoe laces in the normal way. In this case, it seems that whether you have performed a rational action depends on external factors, and not simply on what is going on in your mind. This suggests that the internalist thesis about rationality should not be applied directly to rational actions. Instead, we might say that for an action to be rational is for it to be the execution of a rational intention; and as we have seen, the internalist thesis does apply to rational intentions.

2. Questions about the argument for internalism

The argument for internalism that I have just given rests on an intuitive judgment about these two possible worlds. But should we really trust this intuition? Various questions that might be raised at this point. First, if this intuition is sound, perhaps there are other precisely parallel intuitions that would not just support internalism, but other much more disputable doctrines as well? Secondly, how more exactly are we to draw the line between what is “internal” and “external”? Is there really any line here to which it is plausible to attach any importance? Finally, what could explain why rationality supervenes on what is “internal” in this way? In this section, I shall lay out these questions in greater detail.

The first question concerns whether intuitions of this kind, if they are conceded to be sound, will simply prove too much. The basic point is that the intuition that I have set out above has a certain familiar structure. The intuition concerns two cases, which are radically different in their “external” features, but otherwise as similar as possible with respect to what is going on the relevant thinker’s mind. Considering these two cases is supposed to elicit the intuition that in spite of the radical external differences between the two cases, there is a property of a certain kind that is present in both cases.

This structure is also found in other, more notorious philosophical arguments. First, consider the arguments from *illusion* and from *hallucination* in the philosophy of *perception*. These arguments involve comparing the case of a genuine perception of one’s environment and the case of an illusion or hallucination – two cases that are radically different in their “external” features but as similar as possible in their “internal” features – and then seeks to conclude that there is a property that is shared by these two cases. For example, in certain traditional versions of the argument, such as the version of A. J. Ayer (1973), this common property is called “perceiving a sense datum”.

A second controversial argument that has this structure is an argument for the existence of so-called “narrow content”. Suppose that far away, in a distant galaxy, there is a planet – Twin Earth – that is a perfect duplicate of Earth; and so, one of the inhabitants of Twin Earth is an individual – your Doppelgänger – who is a duplicate of you.⁶ The thoughts of your Doppelgänger seem to have radically different contents from your thoughts. For example, when you are thinking of the central Asian city of Tashkent, your Doppelgänger is thinking of Twin-Tashkent, which is a city on Twin Earth, far away from the city of Tashkent on Earth; and so on. Still, there seem to be some striking and pervasive similarities between your thoughts and your Doppelgänger’s. This has led some philosophers to conclude that there must be a special kind of “narrow content” that is shared by both you and your Doppelgänger.⁷

The argument for internalism that I have just given has a similar structure: it starts with comparing the case of the ordinary thinker with the case of the victim of the evil demon, and concludes that on account of the great “internal” similarities between them, there is an important property – the property of rationality – that they both have in common, in spite of the radical difference in these cases’ “external” features. The main difference between this argument, on the one hand, and these other controversial arguments, on the other hand, is that in the former argument, the alleged common property is a normative property (rationality), while in the other arguments, the alleged common property is a mental property (such as perceiving a sense datum, or having a belief with such-and-such a narrow content).

⁶ The thought experiment of Twin Earth is due to a seminal paper of Putnam (1975).

⁷ For an example of an argument of this sort, see Loar (2003).

This first question that I wish to raise about this argument, then, is this. Can we coherently endorse this argument for internalism about rationality without endorsing those other controversial arguments as well? Will this argument commit us to accepting such controversial ideas as the existence of sense-data and narrow content?

The second question that I wish to explore about this argument for internalism about rationality concerns where exactly the line between what is “internal” and what is “external” should be drawn. Within the epistemological literature, there is dominant approach to drawing this line, in terms of the distinction between the facts that are “accessible” to “reflection alone” and the facts that are not “accessible” in this way. This is sometimes called the “accessibilist” version of internalism. Unfortunately, this approach leads to grave difficulties.

According to this accessibilist approach, the relevant “internal facts”, on which the rationality or irrationality of a belief or process of reasoning supervenes, are defined as “*facts to which one has a special kind of access.*”⁸ One has the relevant “special kind of access” to a fact just in case one is in a position to know that fact “by reflection alone”. In this context, ‘reflection’ means “*a priori* reasoning, introspective awareness of one’s own mental states and one’s memory of knowledge acquired in these ways.” In short, according to this version of internalism, whether or not a belief or process of reasoning is rational “supervenes on facts that one is in a position to know by reflection alone”.

However, it is doubtful whether this accessibilist version of internalism can be true. As Timothy Williamson has argued, there seems not be to any domain of non-trivial facts such that it is a *necessary* feature of all facts within that domain that one is in a position to know those facts by reflection alone.⁹ Thus, the following pair of worlds w_1 and w_2 seems possible. In both w_1 and w_2 , you believe p on the basis of certain reasons, but in w_1 you are in a position to know by reflection alone that you believe p on the basis of those reasons, while in w_2 you are not in a position to know this; otherwise you are in just the same mental states in both w_1 and w_2 .

According to the accessibilist version of internalism, the fact that you believe p on the basis of these reasons may be part of what makes the belief rational in w_1 , but it cannot be part of what makes the belief rational in w_2 . Hence, this version of internalism must allow that it could be the case that this belief is rational in w_1 but not rational in w_2 . But then the sets of facts on which the rationality of the belief supervenes cannot just include the facts about the reasons on the basis of which you hold the belief; every such set of facts must also include some *further* facts, which somehow imply that you are in a position to know about the basis for your belief in w_1 (because – we are assuming – there are no other relevant differences between w_1 and w_2 in this case). According to this version of internalism, whether or not your belief is rational must supervene on some set of facts each of which you are in a position to know by reflection alone; so those further facts must *also* be ones that you are in a position to know by reflection alone. However, according to the point that Williamson has argued for, it is also not a necessary feature of these further facts that you are in a position to know each of them by reflection alone.

In this way, the conjunction of the accessibilist version of internalism with this Williamson-inspired point creates an infinite regress. The facts on which the rationality of your belief supervenes does not just include a set of facts F_0 about the reasons on which the belief is based, and the like. It also includes some further facts F_1 implying that you are in a position to know all members of F_0 by reflection alone; some further further facts F_2 implying that you are in a position to know all members of F_1 by reflection alone; and so on *ad infinitum*. Every

⁸ See Pryor (2001, 103–04).

⁹ See Williamson (2000, Chap. 4).

member of each of these infinitely many sets of facts must be a fact that you are in a position to know by reflection alone. This seems to show that the accessibilist version of internalism cannot be true, at least in anything like the form that we have been considering.

For these reasons then, no true version of internalism can define what it is for a fact to be “internal”, in the sense that concerns them, as a fact that the thinker is “in a position to know by reflection alone”. These “internal facts” must be defined in some other way. But how exactly is the internalist to define these internal facts? If there is no way of defining them, we should perhaps begin to doubt the intuition that lies behind the intuitive argument for internalism.

The final question about this argument that I wish to consider here concerns the *explanation* of *why* internalism is true. Why should be it that there are some genuinely normative concepts that are internalist in this way?

Unfortunately, the main attempt that internalists have made to explain internalism also runs into grave problems. This attempted explanation of internalism is based on the idea that to say that a belief or decision is “rational” is just to say that in holding that belief or making that decision, the thinker is proceeding in a “cognitively blameless” fashion, whereas to say that a belief is “irrational” is to say that the belief is in some way blameworthy or worthy of censure. But it seems that one cannot fairly be blamed for not responding to a fact that one was not in a position to know. So it seems that any facts that are capable of making it the case that that a belief is irrational must be facts that the believer must be in a position to know. This point is held by some philosophers to explain why an internalist conception of rationality must be true.¹⁰

Unfortunately, this attempt at explaining why internalism is true is open to serious objections. First, the claim that rationality is simply a matter of “cognitive blamelessness” seems questionable. There are two ways in which an act can be “blameless” – either because the act was *justified*, or because it was *excusable*. For example, if you kill someone in self-defence, your act may be justified; but if you kill someone because you have gone insane, your act is not justified, although it may be excusable.¹¹ To say that an act is justified implies that the act should be in a way approved of; to say that an act is excusable implies that, although the act should not be approved of, the thinker should not be blamed for having done it. Clearly, the notion of rational or justified belief is much closer to the notion of a justified action than to the more general notion of a blameless action. Thus, not all “cognitively blameless” beliefs are rational or justified: a belief can be blameless merely because it is excusable, even if it is not rational or justified in any way.¹²

Moreover, as Alvin Goldman (1999) has argued, it is doubtful whether the claim that rationality is just a matter of “cognitive blamelessness” does explain this version of internalism. Even if one cannot fairly be blamed for not responding to a fact that one was not in a position to know, it is much less plausible to claim that one can never fairly be blamed for not responding to a fact that one was not in a position to know “by reflection alone”. One can surely be fairly blamed for firing a loaded gun, especially if one thereby hurts or endangers someone. If one was in a position to know whether the gun was loaded or not, one can fairly be blamed for not

¹⁰ For a discussion of the possibility of explaining internalism in this way, see Alston (1989, Essay 8).

¹¹ The distinction between justification and excuse plays an important (albeit contested) role in English and American criminal law; see Corrado (1994).

¹² Essentially this point has been made both by Pryor (2001, 114–18), and by Plantinga (1993, 39). A similar point is also made by McDowell (1994), who distinguishes the “justification” of a belief from mere “exculpation”.

responding to the fact that the gun was loaded. But it is obviously not required that one should be in a position to know that the gun was loaded “by reflection alone”, since that is not the kind of fact that *can* be known by reflection alone; it is the kind of fact that can only be known by perception of one’s environment or by testimony or the like. So this attempt at explaining internalism seems doomed to fail.

A different kind of explanation of internalism is suggested by the argument that John Pollock and Joseph Cruz (1999, 130–40) give against the “reliabilist” claim that the “basic rules” that it is rational for us to conform to in our thinking are those rules that are most reliable at reaching the truth. They understand this as the quite general claim that the rules that it rationally makes sense for us to conform to are *all* and *only* those rules that are reliable in this way – including both rules that we know to be reliable, and rules that we do not know to be reliable in this way.

Pollock and Cruz first argue that if this claim is to address the epistemological issues that concern us, this claim must be a *recommendation* about which rules to conform to. Specifically, it must be the general recommendation that we should conform to all and only reliable rules – in effect, the recommendation to reason in the most reliable way. But they object that this “is not a recommendation anyone could follow”. Their reason is that “we can only alter our reasoning in response to facts about reliability if we are apprised of those facts” (1999, 140).

Here, Pollock and Cruz seem to infer from the premise ‘We can only alter our reasoning in response to facts about reliability if we are apprised of those facts’ to the conclusion ‘No one can follow the recommendation to reason in the most reliable way’. If this inference were valid, then we could also infer from the premise ‘We can only add salt to the water in response to the fact that the water has started boiling if we are apprised of the fact that the water has started boiling’ to the conclusion ‘No one can follow the recommendation to add salt when the water starts boiling’. But that inference *cannot* be valid. Even if the premise is true, it is obviously possible to follow the recommendation to add salt when the water starts boiling.

Pollock and Cruz seem to be assuming that a “recommendation that someone could follow” must be a recommendation that we can *always* follow *whenever* it applies to us. But how many recommendations are there of which that is true? Take the simplest of logical precepts: “From ‘ $p \ \& \ q$ ’ infer p ”. We are not *always* able to follow this recommendation: some conjunctions are too complex to be recognized as such; or, more simply, we might suddenly die, or go insane, or fall asleep, before we have completed the inference; and so on. The only general recommendations that we can always follow, whenever they apply to us, are recommendations that are specifically restricted to cases in which we are able to follow them – for example, “From ‘ $p \ \& \ q$ ’ infer p , whenever you are able to follow this precept”. But recommendations of this sort might be externalist recommendations, such as “Form your beliefs by reliable methods, whenever you are able to follow this recommendation”. So the idea of a recommendation that we are *always* able to follow does not support internalism. So the suggestion that Pollock and Cruz make here also does not seem to provide an explanation of why internalism is true. So it seems that we have not yet identified the correct explanation of why internalism is true.

In the following four sections of this chapter, I shall take each of these three questions about the intuitive argument for internalism in turn: first, I shall discuss its relationship with arguments from hallucination and the like; secondly, I shall investigate how exactly the line between the “internal” and the “external” is to be drawn; and finally, I shall try to see what explanation there might be for why internalism is true.

3. The argument from hallucination and its kin

In the previous section, I listed several questions that one might raise about the intuitive argument for internalism. The first of these questions was based on the thought that the argument has a deep kinship with such controversial arguments as the argument from hallucination. In this section, I shall concede that there is indeed a kinship with those arguments. However, as I shall try to show, when those arguments are stated sufficiently carefully, they are in fact perfectly sound. This point should allay the concern that the intuitive argument for internalism may prove too much: arguments of this kind do indeed have non-trivial consequences, but those consequences are all in fact perfectly true.

The characteristic structure of the arguments that I am concerned with is as follows. First, each of these arguments describes a pair of cases, which are dissimilar in some external respect, but otherwise as similar as they can be in all mental respects, compatibly with their differing in that external respect. For example, such a pair might consist of: (i) a case in which one genuinely perceives one's immediate environment, and (ii) a case in which one has a hallucination that one would not be able to distinguish from such a genuine perception of one's environment. Then, the argument tries to make it plausible that in spite of these external differences, there is a mental property of a certain kind that is present in both cases. Since this mental property is present in both cases despite the difference in the external features of the two cases, this mental state must be independent of these external features.

One of the most famous examples of an argument with this structure is the argument from hallucination. This argument has been widely criticized.¹³ But it seems to me that these criticisms at best undermine certain incautious formulations of the argument, and not the core idea of the argument itself. For example, some incautious formulations of the argument try to conclude, not just that there is a mental property that is present in both of the two cases that the argument focuses on, but that this is a mental state of a very special kind, with a very special object (such as a "sense datum") or a special sort of content (such as a special kind of "narrow content"). As I shall formulate it, the argument from hallucination does not itself try to establish any of these further claims: its conclusion is simply that there is a mental property that is present in both of the two cases, neither more nor less.

Of course, if there is a mental property that is present in both of these two cases, it is natural to ask further questions about this mental property: What sort of mental property is this? And what is the relation between this mental property, which is present in both these two cases, and those mental states that are present in one but not the other of these two cases? However, there is a wide range of answers that could be given to these further questions. While it would indeed be an objection to the argument if there were *no* plausible answer that could be given to those further questions, the argument itself is not tied to any specific answer to those further questions.

To fix ideas, let us take the pair of cases that Mark Johnston (2004, 122) invokes in his statement of the argument from hallucination. You are undergoing brain surgery, while quite conscious, under local anaesthetic. The surgeon "applies electrical stimulation to a well-chosen point on your visual cortex. As a result, you hallucinate dimly illuminated spotlights in a ceiling above you. ... As it happens, there really are spotlights in the ceiling at precisely the places where you hallucinate lights." Then: "the surgeon stops stimulating your brain. You now

¹³ For such criticisms, see Dancy (1995) and McDowell (1994). For a defence of the argument, see Johnston (2004).

genuinely see the dimly lit spotlights in the ceiling. From your vantage point there on the operating table these dim lights are indistinguishable from the dim lights you were hallucinating. The transition from ... hallucination to ... veridical perception could be experientially seamless. Try as you might, you would not notice any difference, however closely you attend to your visual experience.”¹⁴

Some philosophers may think that they can directly intuit that there is a mental property that is shared between these two cases. This would make the argument from hallucination strictly analogous to the argument for internalism about rationality that I am considering here – since we seem to intuit that there is an important normative property (which we can stipulate to be referred to by our use of ‘rationality’) that is present in both the two cases that that argument focuses on. With the argument from hallucination, however, some philosophers will deny that we can intuit that there is any such common mental property. However, there is an argument – taking the form of an *inference to the best explanation* – for the conclusion that there must in fact be such a common property.

What does it mean to say that “from your vantage point”, the dim lights that you see in the ceiling are “indistinguishable from the dim lights you were hallucinating”? It seems to mean this: you lack any reliable ability to respond to the genuine perception by forming different beliefs and judgments from the beliefs and judgments that you would form in response to the hallucination. This is because in each of these two cases, you are disposed to form almost exactly the same beliefs and judgments (and the same doubts and uncertainties) about what is going on in your environment, about your own mental states, and so on.¹⁵ This is why you fail to notice any change in your circumstances when you switch from hallucinating to genuinely perceiving the lights. In general, the two cases seem remarkably similar in their causal powers – at least in their immediate short-term causal powers to cause other mental states and events.

What can explain this remarkable fact that these two cases are so extraordinarily similar with respect to the mental states and events that these cases are disposed to cause? This is puzzling – especially because in Johnston’s version of the argument, you switch from the *bad* case of hallucination to the *good* case of genuinely perceiving. Since your cognitive situation has improved, it is surprising that you do not notice how it has changed.

The best explanation seems to be that there is a mental property that is present in both of these two cases, and it is this common mental property that disposes you to form all the same beliefs and judgments in both cases. As I noted above, I do not have to take a definite stand on the further question of what exactly this common mental property is. Many different answers to this further question are possible. For example, one possible answer is that in this pair of cases, the common mental property common is the property of *having an experience as of there being dimly illuminated lights in a ceiling above you*. But the crucial point is that the best explanation is that there is a common mental property in both cases.

Some philosophers deny that there is any mental state that is common to the two cases. According to these philosophers, the two cases involve fundamentally different mental states – in

¹⁴ Johnston actually focuses on three cases: a hallucination whose content is false or non-veridical, a veridical hallucination, and a genuine perception. It seems to me however that this additional complication is not strictly necessary for the argument.

¹⁵ I say “almost exactly the same beliefs and judgments” because strictly speaking demonstrative judgments (such as the judgment that *those lights there* are dim) will be different in the two cases, as we can see from the fact that such demonstrative judgments will have different truth conditions in the two cases.

the one case a hallucination, and in the other a genuine perception; all that these cases have in common is that both cases involve the *disjunction* of these two mental states – that is, they both involve the disjunctive state of *either* hallucinating spotlights in a ceiling *or* seeing spotlights in the ceiling.¹⁶

However, this “disjunctivist” response fails to provide any explanation of something that cries out for explanation – namely, how it can be that these two cases are so similar with respect to the beliefs and judgments that one is disposed to form in those cases. After all, *any* two cases in a thinker’s mental life, no matter how dissimilar these cases may be from each other, will both involve the disjunction of some mental state involved in the first case and some mental state involved in the second. For example, consider one case in which I am in excruciating agony, and another in which I am listening to some beautiful music. These two cases have in common that they both involve the disjunctive state of *either* being in excruciating agony *or* listening to some beautiful music. But that the two cases have this much in common would hardly explain any other similarity that they might have (such as a striking similarity in the beliefs and judgments that one is disposed to form in those cases).

Another alternative suggestion that might be made is that the similarities between the two cases can be explained if there are *neural* properties that are shared between these two cases; according to this suggestion, there is no need to postulate any shared *mental* property. However, the similarities between the two cases are not merely a surprising empirical discovery that we have made. Relying only on the grasp of folk psychology that we all have, in virtue of being competent users of the folk-psychological concepts (like *belief* and *intention* and the like), we find it intuitively intelligible that these similarities would exist between these two cases. This seems to show that the properties that are shared between these two cases are properties of the sort that are recognized, at least implicitly, in ordinary folk-psychological thinking – in other words, the shared properties must be genuine mental properties.

In fact, there is another version of this sort of argument, which supports the conclusion that there is a mental property that is common to the case of *knowledge* and the case of *false belief*. Consider one case where a thinker knows that Tashkent is the capital of Uzbekistan, and another case where the thinker has a false belief that Tashkent is the capital of Uzbekistan. (Suppose that this second case obtains in a different possible world, where the capital of Uzbekistan is not Tashkent but Samarkand instead.) In spite of this difference between the two cases, there seems to be a striking and important similarity in the causal powers of the thinker’s mental states in these two cases (including their causal powers to lead the thinker to have introspective beliefs about her own mental states). So it seems that we should reject a purely disjunctive conception of what the two cases have in common;¹⁷ instead, we should conclude that there is a genuine mental property in common – presumably, the property of *believing* that Tashkent is the capital of Uzbekistan.

Another argument that is an instance of the general type that we are considering here focuses on the example of Earth and Twin Earth. Suppose that you are transported from Earth to Twin Earth in your sleep, and that you then remain on Twin Earth for the rest of your life. At

¹⁶ This is the view of the “disjunctivist” school of thought. For some canonical statements of this disjunctivism, see Hinton (1973), Snowdon (1981), and McDowell (1994). For criticism of some of the arguments that have been used to support disjunctivism, see Millar (1996).

¹⁷ Compare Williamson’s (2000, 44–46) argument that we should not adopt a disjunctive definition of belief, according to which believing *p* is the disjunctive state of either knowing *p* or merely opining *p*.

some point, it will be Twin Earth, rather than Earth, that counts as your normal environment, and it will be a community on Twin Earth, rather than any community on Earth, that counts as your community. At that point, then, it seems, your terms and concepts will switch from referring to the objects and kinds of Earth to referring to the objects and kinds of Twin Earth. But it is striking that you do not notice any switch in the content of your thoughts. This change seems to leave everything else about your mental states and dispositions unchanged. But that is an extraordinary fact. How can the contents of all your thoughts change so thoroughly and yet leave so much intact? You might even move back and forth between Earth and Twin Earth several times, in which case the contents of your thoughts might change back and forth several times. How is it possible for such repeated cognitive revolutions to escape your attention?

The best explanation of this, it seems to me, is that there is an explanatorily significant mental property that is common to both the Earth case and the Twin Earth case. In saying that there is a “mental property” present in both cases, I am not requiring that this mental property must take the form of standing in a specific mental relation to a particular content. Again, I do not need to take a definite stand on the further question of what exactly this common mental property is. But as I shall now explain, it seems to me that there is a plausible answer that can be given to this further question.

In particular, to answer this further question we do not need to appeal to the controversial idea of “narrow content”. As it is normally understood, the “content” of a thought necessarily involves the thought’s truth conditions. But almost all our thoughts are about the external world; in that sense, almost all our thoughts have truth conditions concerning the external world. So we may conclude that the only kind of content that the overwhelming majority of our thoughts have is “broad content”. That is, the intentional content of almost all thoughts depends on the thinker’s relations to her environment.

However, it may still be possible to classify these broad contents into certain *narrow types*. In effect, we can sort these broad contents into types, or equivalence classes, such that purely internal facts about the thinker are enough to determine that she is believing a content of *such-and-such a type*, even though it is not enough to determine precisely *which* content of this type she is believing. For example, it may be that for a content to be of such-and-such a narrow type is for it to have such-and-such a logical form, and to be composed out of *concepts* of such-and-such narrow types; and the relevant narrow types of concepts might be determined by what we could call the “internal conceptual role” of concepts, as opposed to their environmentally determined reference or semantic value. Strictly, however, it does not matter for my purposes exactly how these narrow types of content are defined – only that contents can be classified into narrow types of this kind. Once we have identified the narrow type of content *T* that a particular broad content belongs to, we can identify the mental property that is shared both by the thinker on Earth (who believes this particular content) and by that thinker’s Doppelgänger on Twin Earth (who believes a different content of the same type): a thinker *x* has this mental property if and only if there is some content *y* that belongs to this type *T* such that the thinker *x* believes *y*.

At all events, this is just a suggestion about how we might identify the mental property that is shared between the cases of Earth and Twin Earth. The main argument that I have given does not depend in any way on this precise identification of the shared property. All that this argument implies is that there is some explanatorily significant mental property or other that is shared between the two cases.

The conclusion of each of these arguments, then, is that (a) there is a mental property that is common to both cases, in the relevant pair of cases, and (b) since this property is present in

both cases, it must be independent of the external factors that vary between the two cases. To gauge the significance of this conclusion accurately, it is important to remember how these pairs of cases have been specified. They have *not* been specified simply as pairs of cases in which everything that is “inside the head” is the same, but what lies “outside the head” is different. On the contrary, they have been specified as pairs of cases in which certain *specific* external features are different (for example, in the first pair, one case involves genuine perception, while the other involves a hallucination), while otherwise the two cases are stipulated to be as similar as possible in all *mental* respects. The overall goal of the argument is to draw attention to the existence of genuine mental properties that are independent of the specific external factors that are specified as differing between the two cases.

In this way, each of these arguments only establishes that the mental property common to the two cases in question is independent of the *specific* external factor that differs between the two cases. These arguments do *not* establish that the mental property in question is independent of everything that is “outside the head”; indeed, these arguments are compatible with the thesis that *no* mental properties are independent of everything that is going on outside the head.¹⁸ Much further investigation would be required to determine whether or not these mental properties depend on *some* factors that lie outside the head. Each of the arguments surveyed in this section implies only that these mental properties are independent of the specific external factor that is specified as differing between the two cases.

When these arguments are understood in this way, it seems plausible to me that these arguments are sound. Admittedly, in each pair of cases, we can the two further questions that I mentioned earlier. First, what exactly are the internal mental properties that are shared between the two cases? Secondly, how exactly are these internal properties related to the “broad” mental states that differ between the two cases? As I noted above, the conclusion of each of these arguments is simply that there is a mental property common to both cases; and this conclusion does not depend on the correctness of any particular answers to these further questions. But to fix ideas, it may be helpful to suggest some possible answers to these further questions.

In answer to the first of these further questions, I have already made suggestions about what these internal mental properties are. Relative to the difference between knowing *p* and having a false belief in *p*, the internal mental property is *believing p*; relative to the difference between a genuine perception and a hallucination, the internal property is the state of *having an experience as of p's being the case*; relative to the difference between Earth and Twin Earth, the internal property is the property of having an attitude of the relevant kind towards some (broad) content or other *of the narrow type T*.

What about the second of these further questions? For example, what is the relationship between the broad state of knowing *p* and the more internal property of believing *p*? What is relationship between the state of believing *p* and the yet more internal property of believing a proposition of narrow type *T* (where the proposition *p* is in fact of type *T*)? In both cases, the

¹⁸ So, for example, it is quite compatible with the conclusions of these arguments that a brain in a vat that had *never* been connected to a body that was capable of acting in a normal environment would be totally incapable of having any mental states at all. If so, then none of these “internal” states will supervene purely on intrinsic features of the thinker’s brain; they will supervene only on a *slightly wider* supervenience basis, which might include certain highly general and unspecific features of the thinker’s environment. Nonetheless, the supervenience basis for these internal states would still be much narrower than that of factive states like knowing or perceiving that something is the case.

relationship is clearly one of one-way strict implication: necessarily, if one is in the broad state of knowing p , then one is in the relatively internal state of believing p ; and necessarily, if one is in the state of believing p , then one has the yet more internal property of believing a proposition of type T ; and in both cases, the converse implication fails to hold. This makes it plausible that the relationship is that of a *determinate* to a *determinable*, as the property of being scarlet is a determinate of the determinable property of being red, and the property of being an equilateral triangle is a determinate of the determinable property of being a triangle. Thus, for example, the state of knowing p is a determinate of the determinable property of believing p , which is in turn a determinate of the yet more determinable property of believing some proposition or other of the narrow type T (where p is a proposition of this type T).

As I have emphasized, the soundness of the arguments that we have been considering here does not depend on any particular answers to these further questions. Still, the fact that there seem to be plausible ways of answering these further questions should raise our confidence that these arguments are indeed sound. The intuitive argument for internalism about rationality sketched in Section 1 above is akin to the other arguments that we have explored in this section; the main way in which it differs from those other arguments is that while each of those other arguments seeks to show that there is a *mental* property that is present in both of the relevant cases, the argument for internalism seeks to show that certain *normative* features – like rationality and irrationality – are present in both cases. In general, however, this argumentative strategy seems to me perfectly sound. The fact that the intuitive argument for internalism is akin to controversial arguments like the argument from hallucination does not ground any objection to the argument at all.

4. The internal / external distinction

As I shall argue in this section, this interpretation of this family of arguments helps to clarify how internalists about rationality should draw the distinction between what is “internal” and what is “external” in the relevant sense.

As I have explained, each of these arguments focuses on a pair of cases. In each of these pairs of cases, it is assumed that the two cases differ in some respect that concerns the relationship that the thinker has to certain factors that lie outside the thinker’s head in the thinker’s environment. For example, in the argument from hallucination, the two cases differ in the kind of causal relationship that holds between the thinker’s mind and the scene before the thinker’s eyes: in one case, there is the sort of causal connection between the thinker’s mind and her environment that is involved in the thinker’s genuinely *perceiving* her environment, while in the other case, this connection is absent. In the case involving Earth and Twin Earth, there is a relationship between the thinker’s mind and Earth in the first case, and an analogous relationship between the thinker’s mind and Twin Earth in the second case.

What each of these arguments shows is that there are mental properties that are independent of whether or not this specific relationship holds between the thinker’s mind and her wider environment. In this way, these arguments do not presuppose any *absolute* distinction between the “internal” and the “external” at all. Instead, we may take these cases to define a *relative* distinction between the “internal” and the “external”: if C is an external condition – that is, a condition that concerns whether or not a certain relationship holds between the thinker and what is going on outside the thinker’s head – then any mental properties that can remain

unchanged while C varies are mental properties that are independent of this external condition C , and so in effect internal relative to C .

In general, we can recognize what we might call “degrees of internality”. Specifically, we could say that one fact is “more internal” than a second if and only if the external conditions that the first fact is independent of *properly include* the external conditions that the second is independent of.

In practice, there seem to be three main degrees of internality. First, some mental states, like *knowledge*, are typically dependent on the external world *both* for the type of *attitude* that they involve *and* for their propositional content. These states are dependent on the external world for the attitude that they involve because the difference between knowing a proposition p and believing a proposition p depends in part on whether p is true – which is typically a matter of how things are in the external world. These mental states depend on the external world for their content because I know that Tashkent is the capital of Uzbekistan, while the content of my Twin-Earth Doppelgänger’s knowledge is different – what he knows is the quite different fact that Twin Tashkent is the capital of Twin Uzbekistan. These mental states are the least internal of all.

Secondly, some mental states, like *beliefs* and *experiences*, are typically dependent on the external world for their content, but *not* for the type of attitude that they involve. These states do not depend on the external world for the type of attitude that they involve, because both I and my Twin Earth Doppelgänger have beliefs and experiences – and similarly, you have beliefs and experiences both in world w_1 , where you are bedevilled by an evil demon, and in world w_2 , where your experiences are largely reliable and your beliefs are largely true. But beliefs and experiences depend on the external world for their content, because I and my Twin Earth Doppelgänger have beliefs with different contents (for example, my beliefs concern Tashkent, while his beliefs concern Twin Tashkent). These mental states are internal to an intermediate degree.¹⁹

Finally, some mental states, like states of *believing some content or other of the narrow, internally-individuated type T* , are dependent on specific relationships with the external world *neither* for their attitude-type *nor* for their content. These mental states are present in both cases in all the pairs of cases that we have considered. Both in the evil-demon world and in the normal-perception world, and both on Earth and on Twin Earth, the same mental states of these kinds are present. These mental states are internal to the highest degree.

This helps to clarify the kind of “internal” facts about the thinker’s mind that determine the rationality or irrationality of her beliefs. In Section 2, we saw that one standard way in which internalism has been formulated – as the claim that rationality supervenes on facts that one is in a position to know “by reflection alone” – runs into fatal problems. As Earl Conee and Richard Feldman (2004) have put it, there are two main kinds of internalism: one kind – accessibilism – claims that the rationality or irrationality of a belief supervenes on facts that are “accessible” in a certain special way; the other kind – which they call “mentalism” – claims that the rationality or irrationality of a belief supervenes on facts that in some way or other concern the thinker’s

¹⁹ In principle, there are also mental states that have an intermediate degree of internality in a different way – by being states that depend on the environment for their attitude-type but *not* for their content. The state of *knowing some content or other of type T* would be an example. Unlike states (like *believing p*) that depend on the environment for their content but not for their attitude-type, these states do not play a prominent role in ordinary folk-psychological thought; I shall ignore these mental states here.

mental states. The arguments that I gave in Section 2 show that accessibilism has to be rejected. The kind of internalism that I am defending here is of the mentalist (not the accessibilist) variety.

If I am right that accessibilism must be abandoned in favour of mentalism, then what results from the intuitive argument of Section 1 is a quite different form of internalism from the forms that many other philosophers have defended.²⁰ The fundamental idea behind the accessibilist form of internalism is that what makes a belief rational or irrational must always be something that the thinker herself is in a position to know about. Rationality and irrationality depend on what sort of higher-order perspective on one's own thinking is available. Since we only rarely actually form such higher-order beliefs about our thought-processes, the accessibilist form of internalism implies that rationality of our actual thinking depends on the higher-order beliefs that we *would* form about that thinking, under certain possible but usually non-actual circumstances. The mentalist form of internalism makes the quite different claim that the rationality of our thinking depends on the *actual* character of that thinking itself, not on facts about what we would believe about that thinking.

However, just saying that the intuitive argument of Section 1 supports a kind of mentalism does not yet help us to see what facts count as "internal facts" in the relevant sense. The problem is that not all mental states are "internal" in this sense. As Timothy Williamson (2000) has argued, there is no reason not to include *knowledge* as a mental state, at least of a kind. But knowledge, as I have explained, is one of the least internal of all mental states. This is why all philosophers who describe themselves as "internalists" about rationality insist that the rationality of a belief is independent of the facts about what the thinker knows. But if the relevant "internal" mental states do not include knowledge, then how exactly are they to be specified?

In fact, on closer inspection, it becomes clear that the question of what the rationality of an attitude supervenes on can be taken in two ways. If it is rational for a thinker in a certain situation to have a certain attitude, there is a kind of *rationalizing relation* that holds between the thinker's situation and the attitude in question. When the question about what the rationality of an attitude supervenes on is taken in the first way, we focus on a particular attitude, and inquire about the property of being a situation that has the rationalizing relation to this particular attitude. For example, we might inquire about the property of being a situation that has this rationalizing relation to the attitude of believing a particular proposition *p*. By contrast, when the question about what rationality supervenes on is taken in the second way, we focus on the *pair* consisting of the situation and the attitude in question, and inquire about the property of being a pair consisting of a situation and an attitude such that the situation has this rationalizing relation to the attitude.

Taken in the first way, the question concerns the property of being in a situation in which it is rational to believe the particular proposition *p*. For example, suppose that *p* is the proposition that Tashkent is the capital of Uzbekistan. It seems that it is rational for me to believe this proposition *p*, but not rational for my Doppelgänger on Twin Earth to believe this proposition. (It is rational for him to believe a different proposition – the proposition that Twin Tashkent is the capital of Twin Uzbekistan – instead.) So, it seems, the property of being in a situation in which it is rational to believe *p* supervenes on facts about the thinker's mental states and events that are internal to an *intermediate* degree – less internal than facts about "factive" mental states like knowledge, but not so internal that these facts do not vary between Earth and Twin Earth.

²⁰ For some prominent defenders of accessibilist forms of internalism, see for example Fumerton (2001) and BonJour (2001).

Taken in the second way, however, the question concerns the property of being a *pair* consisting of a situation and an attitude such that the situation has the “rationalizing relation” towards the attitude. This property does not involve any particular proposition *p*. It seems that a situation-attitude pair’s having this property supervenes on facts that are internal to the *highest* of the three degrees that I characterized above. Suppose that w_1 and w_2 differ not only in the fact that in w_1 you are bedevilled by an evil demon, whereas in w_2 your experiences are reliable perceptions of your environment, but also in the fact that in w_1 , you were transported to Twin Earth while you were an infant, whereas in w_2 , you remained on Earth. It still seems that every belief that you have in either of these two worlds can be paired off with a corresponding belief that you have in the other world, in such a way that (a) the corresponding beliefs share all the mental properties that are internal to the highest degree, and (b) in each case, each belief is just as rational, or as irrational (in the situation in which it is held) as the corresponding belief in the other case (in the situation in which it is held).

What this seems to show is that the property of being a situation-attitude pair such that in the situation it is rational to have the attitude supervenes on mental facts that are internal to the highest degree. Suppose that in the thinker’s current situation, it is rational for the thinker to believe *p*. Then any situation-attitude pair that resembles this one in all maximally internal respects will also resemble it with respect to consisting of a situation and an attitude such that in the situation, it is rational for the thinker to have the attitude.

This, then, is the answer that I propose to the question of how to draw the line between what is “internal” and what is “external” for the purposes of understanding internalism about rationality. However, we still need to know why this version of the mentalist form of internalism is true. This form of internalism is clearly quite different from the more familiar accessibilist form. So presumably its explanation would also have to be quite different from anything that could explain the truth of any accessibilist form of internalism. We shall try to see how this form of internalism could be explained in the next two sections.

5. Direct guidance

Why should rationality supervene purely on internal mental facts in this way? Let us focus again on what in the previous section I called the “rationalizing relation”, which holds between a thinker’s situation and an attitude (like a belief) if and only if it is rational for the thinker to have that attitude in that situation. This relation – I propose – is a normative relation of a special kind: the facts about the attitudes and ways of thinking that are rational in my situation must be *capable of directly guiding* my thinking in this situation. In this section, I shall explain what I mean by talking about a fact’s being “capable” of “directly guiding” someone’s thinking.

To make the discussion more concrete, I shall focus on a specific instance of what it is to be guided by a fact in one’s thinking – specifically, I shall focus on cases in which the fact that one’s situation has a certain feature guides one towards *forming* a certain *belief*. Presumably, facts can guide one’s thinking in other ways too – such as the way in which one revises or maintains one’s beliefs, or the way in which one forms, maintains, or revises other mental states besides belief. The account that I shall give of what it is for a fact to guide one’s thinking is designed to be capable of being generalized to apply to those other cases, but I shall not work out the details here.

Intuitively, one might form a belief that it is rational for one to form – through sheer dumb luck. It might be that it is rational for you to form a belief in *p*, and as a matter of fact you

do form that belief – but it is a complete fluke that you form that belief at the same time as its being rational for you to do so. In this case, you form a certain belief in a situation that stands in the rationalizing relation to forming that belief; but you are not in any way *guided by* the fact that your situation rationalizes your forming that belief.

To capture what is involved in being guided by the facts about what beliefs it is rational to form in one's situation, it would not be enough just to say that one must form a belief that it is rational in one's situation directly in response to being in that situation. Suppose that whenever you consider a proposition of the form ' $\neg(p \ \& \ \neg p)$ ', it is rational for you to form a belief in that proposition. You might consider a proposition of this form, and form a belief in the proposition in direct response to your considering the proposition. However, it could also be that the fact that you form a belief in this proposition has nothing to do with the fact that it is rational for you to do so. For example, a devious neuroscientist might have manipulated your brain so that you would form a belief in *whatever* proposition you considered at that time, regardless of whether or not it was a logical truth or not. In this case too, it seems to me, it is a fluke that you are forming a belief that it is rational for you to form; you are still not really being guided by the fact that it is rational for you to form the belief. So, I propose, for you to be guided by the fact that you are in a situation in which it is rational for you to form a belief in this proposition, it must be *no accident* that you form this belief in a situation in which it is rational for you to do so.

What is it for it to be no accident that you form this belief in a situation in which doing so is rational? I propose a *dispositionalist* answer to this question.²¹ For this to be no accident, you must form the belief through manifesting a disposition of the right sort. For every disposition of the right sort, there must be some range of situations $S_1, \dots S_n$, and some relation R that can hold between each of these situations and a proposition, such that in each situation S_i it is rational to form a belief in any proposition that stands in relation R to S_i ; and the disposition must be a general disposition to respond to being in a situation in this range by forming a belief in a proposition that stands in relation R to that situation. For you to be “guided”, in forming a certain belief, by the fact that it is rational for you in your situation to form this belief, is simply to manifest a disposition of this sort.

When you manifest a disposition of this sort, it is precisely *because* you are in a situation that stands in this relation to the proposition in question that you form a belief in that proposition. The fact that you are in a situation of this sort is at least part of a correct causal explanation of the fact that you form the belief. (I shall assume here that the entities that enter into explanatory relations either as the thing that gets explained – the *explanandum* – or as the thing that does the explaining – the *explanans* – are *facts*, rather than events.)

This, then, is what I mean by saying that you are being “guided by” the fact that it is rational for you to have a certain attitude. But what do I mean by saying that you are “*directly*” guided by this fact? The basic idea is quite simple: to be directly guided by a certain fact, you must be guided by that fact, but *not* by means of any further process of reasoning that is *distinct* from the process of being guided by that fact. Being guided by a fact “directly”, in this sense, is analogous to performing a *basic action*. A basic action is an action that one performs, but not by means of performing any other action.²² In a similar way, being guided by a fact directly is being guided by that fact, but not by means of any further process of reasoning that is distinct from being guided by that fact.

²¹ For a more detailed discussion of this dispositionalist approach, see Wedgwood (2006) and (2014).

²² The notion of a basic action is due to Arthur Danto (1968).

What does it mean to talk of a certain fact's guiding one's thinking "by means of" some distinct piece of reasoning? There are two main kinds of case that I have in mind here.

First, you might perform one piece of rational reasoning by means of performing a number of other simpler pieces of rational reasoning. For example, you might form a belief in a theorem by means of considering a long and complex proof. In this case, the fact that it is rational for you to form a belief in the theorem by means of considering the proof does in a sense guide you towards forming that belief; and your being guided by this fact is a process of reasoning. However, this fact guides you by means of a host of *other* simpler pieces of reasoning – such as simple pieces of reasoning in which you accept inferences that have the form of *modus ponens* and the like.

In cases of this kind, the process whereby you consider the proof and thereby form a belief in the theorem can itself be analysed, even at the folk-psychological level of explanation, into a series of *sub-processes* or *steps*, each of which is a simple piece of reasoning. Ultimately, however, this kind of analysis comes to an end, when we have analysed the process into a series of steps that cannot themselves be analysed, at least at the folk-psychological level of explanation, into any more basic sub-processes. These are the basic steps in your reasoning out of which the whole of your reasoning process is composed. If you are guided by any fact in your performing these basic steps, then that fact would guide you directly, not by means of any other simpler pieces of reasoning.

There is also a second kind of case that we should take note of. In cases of this second kind, a fact guides you by means of a process which has some proper parts that are processes of reasoning, but also other parts that are *not* processes of reasoning. For example, in some cases, you might be guided towards forming a belief in p by the fact that p is true. However (except perhaps in cases where the truth of p is directly accessible to introspection), if you are guided by the truth of p , this truth will guide you at least partly by means of a process of reasoning that is not itself identical to the process of your being guided by the truth of p . For instance, the truth of p might guide you towards forming a belief in p by means of a process that includes your forming this belief in response to a sensory experience in which it appears to you that p is the case – and that process is a simple piece of reasoning. At the same time, this process leading from the experience to forming this belief is not identical to the process whereby the truth of p itself guides you towards forming this belief, since the latter process also includes another part – namely, a process whereby the fact of p 's being the case causes you to have a sensory experience of this sort. So the process whereby the fact that p is true guides you towards forming a belief in p includes a distinct sub-process that is itself a process of reasoning. For this reason, the fact that p is true does not directly guide you in this case – the guidance that this fact provides is mediated by this simple piece of reasoning.

In this way, in these two kinds of case, one's reasoning is guided by a fact, but only by means of a complex process that itself involves a distinct process of reasoning as a sub-process; in such cases, the fact does not guide one's reasoning directly. By contrast, when a fact guides one's reasoning directly, the process of one's being guided by that fact cannot be analysed, at the folk-psychological level of explanation, into a series of sub-processes that include any other process of reasoning. In general, it seems plausible that if a process of this sort could be analysed, at the folk-psychological level of explanation, into any series of sub-processes at all, then at least some of those sub-processes would have to be processes of reasoning. So, if a fact guides one's reasoning directly, the process whereby it guides one's reasoning cannot be analysed, at least at the folk-psychological level of explanation, into any sub-processes at all.

There are no intervening steps that can be captured at this folk-psychological level of explanation at all.

When a correct folk-psychological explanation of one's forming a belief includes all the intervening steps that can be captured at the folk-psychological level, we may call it a "fully-articulated" explanation. In general, if one is directly guided towards forming a certain belief by a certain fact, then such a fully-articulated explanation of your forming that belief will identify that fact, or some further fact on which that fact supervenes, as (at least part of) the *proximate explanation* of your forming that belief. According to the proposal that I am making here, then, the fact that it is rational for you to form a belief – or some further fact that the fact in question supervenes on – must be capable of being the proximate explanation (at the folk-psychological level of explanation) of your actually forming that belief.

It is important that this claim only concerns the *personal, folk-psychological* level of explanation. At a "subpersonal" level of explanation, it may well be that the process of one's being directly guided by such a fact can be analysed into numerous sub-processes, perhaps involving various subpersonal modules' computing various algorithms. But this is not the sort of explanation that we are concerned with here. We are concerned with folk-psychological explanations – or more precisely, explanations that have the following two features.

First, these explanations are at the *personal, mental* level: what is explained is a mental fact about a person as a whole – such as the person's having or forming a certain mental state, like a belief or an intention, of the sort that are referred to in everyday folk-psychological discourse. Moreover, this fact is explained by reference to other states of the person as a whole; these explanations do not refer to states of subpersonal mechanisms or regions of the brain or anything of that sort.

Secondly, these explanations make the person's having or forming that mental state *intuitively intelligible* or *unsurprising*. For example, the fact that John decided to go to the florist's shop this morning is made intuitively intelligible by the fact that he wanted to buy some flowers, and believed that the best way to do this was to go to the florist's shop this morning. On the other hand, John's going to the florist's shop this morning is *not* made intuitively intelligible by the fact that he wanted to see the new Steven Spielberg movie, and believed that the best way to do that is to go to the cinema in the evening.²³ It is explanations that have these two features that I am referring to as "folk-psychological explanations".

My proposal, then, is that the rationalizing relation is a distinctive sort of normative relation: the facts about the beliefs it is rational for you to form in your situation must be capable of in this sense directly guiding you towards forming those beliefs in that situation. More precisely, this proposal has two parts. First, the fact that it is rational for you in your situation to form a belief in the particular proposition *p* must be capable of directly guiding you towards forming a belief in *p*. Secondly, when it is rational for you to form a belief in a certain proposition, that is because the proposition stands in a certain relation to your situation, and the fact that your

²³ There are many theories about what it is to make someone's having or forming a certain mental state "intuitively intelligible" in this way. On some theories, it is a matter of explaining the mental state in accordance with a certain tacitly known folk-psychological *theory*. On other theories, it is a matter of *Verstehen* – that is, imaginative projection into, or simulation of, the person's point of view. I will remain neutral between these different theories here. I shall simply have to rely on the reader's having an intuitive sense of when explanations succeed in making someone's having or forming a certain mental state intuitively intelligible or unsurprising.

situation has that relation to a proposition must be capable of directly guiding you towards forming a belief in a proposition that stands in this relation to your situation.

What do I mean by proposing that such facts are “capable” of directly guiding you towards forming such beliefs? What I have in mind is stronger than the proposal that that it is barely metaphysically or logically possible that this fact will directly guide you in this way. Specifically, I mean that in *normal circumstances*, there is a *significant* chance that this fact will directly guide you – that is, that this fact, or some further fact on which it supervenes, will be at least part of the proximate explanation of your forming such a belief.

The kind of “normality” relevant here is the kind that is presupposed by standard folk-psychological explanations. It seems plausible to me that folk psychology involves certain general principles or causal regularities.²⁴ Normal circumstances are precisely those circumstances in which things transpire in accordance with these regularities. So, if the relevantly normal circumstances involve a significant chance of a certain fact’s being the proximate explanation of your forming a certain belief, these causal regularities will assign a significant chance to that fact’s being the proximate explanation of your forming that belief.²⁵

This, then, is what I mean by my proposal that the facts about how it is rational for you to think in your situation must be capable of directly guiding your thinking in that situation. In the following section, I shall argue that this proposal explains why the internalist conception of rationality is true.

6. Why internalism is true

In the previous section, I made a proposal about what is distinctive of rationality, when compared with other normative properties and relations. According to this proposal, the distinguishing feature of rationality is that the facts about the way in which it is rational for a thinker to think in a given situation are, as I put it, “capable” of “directly guiding” that thinker in that situation. I illustrated this proposal by explaining what it is for the fact that it is rational for you to form a certain belief in your current situation to “directly guide” you towards forming such a belief in that situation; I also explained what it is for such facts to be “capable” of providing such direct guidance. The crucial question concerns the structure of the general causal regularities of folk psychology: according to these regularities, what kind of facts have a significant chance of being the proximate explanation of the thinker’s forming a certain belief in normal circumstances?

As I shall argue here, if the thinker’s circumstances are normal in the relevant way, the only facts that – according to the general principles or regularities of folk psychology – have a significant chance of being the proximate explanation of the thinker’s forming a belief are facts

²⁴ For a compelling argument for this point, see Antony (1995).

²⁵ It is not necessary that circumstances will be normal in this sense, and so it is possible for this fact to obtain even though there is no significant chance that the fact in question will be the proximate explanation of your forming a belief. Still, it is arguably impossible for this fact to obtain without its being the case that *in normal circumstances*, there is a significant chance of its being the proximate explanation of your forming a belief. Since this is what I mean by saying that the fact is “capable of directly guiding you”, it is arguably a necessary feature of this fact that it is capable of directly guiding you. For this reason, my proposal does not face an analogue of the vicious regress that I raised in Section 2 against the accessibilist version of internalism.

about the thinker's internal mental events and states.²⁶ I shall argue for this in two stages. First, I shall argue that in circumstances of this kind, the proximate folk-psychological explanation of a thinker's forming a belief is always some fact about that thinker's *mental* events and states. Then I shall argue, more specifically, that in such circumstances, the proximate explanation of the thinker's forming a belief is always an *internal* fact about these mental events and states.

In arguing for the first point, I am not denying that it is *ever* correct to explain why a thinker forms a belief, in normal circumstances of this kind, on the basis of something other than a fact about the thinker's mental events and states. For example, it might be correct to explain my forming the belief that Fermat's last theorem is true, even in perfectly normal circumstances, on the basis of the fact that I have been *told by a reliable informant* that Fermat's last theorem is true – even though the fact that I have been told by a reliable informant that Fermat's last theorem is true is not a fact about my mental events and states. This explanation may be quite correct. It just does not identify the *proximate* psychological explanation of my forming the belief that Fermat's last theorem is true.

Intuitively, it seems, if this is a correct explanation, there must also be a more *detailed* correct explanation, in which my forming the belief that the theorem is true is not directly explained by my being told by a reliable informant that Fermat's last theorem is true, but is instead explained by some intervening fact about my mental states. For example, perhaps my forming the belief that Fermat's last theorem is true is explained by my having the *belief* that I have been told by a reliable informant that the theorem is true; and my having this belief (that I have been told by a reliable informant that the theorem is true) is itself explained by my having an *experience* as of someone (whom I take to be a reliable informant) telling me that the theorem is true.

Suppose that I claim that a thinker's forming a certain belief is explained by a certain external fact that is not a fact about that thinker's mental states; and suppose that the context does nothing to make it clear how there could be any more detailed correct explanation in which the link between that external fact and the formation of that belief is mediated by any intervening facts about the thinker's mental states. For example, suppose that I say, "I once lived in Edinburgh, so Barack Obama formed the belief that I once lived in Edinburgh". It would be natural for you to reply: "But how does Obama know anything about you at all? Did you meet him and talk about your life? Did he have you investigated you by the CIA? Or what?" In asking these questions, you reveal that you would not accept this explanation unless it is plausible to you that this link, between the fact that I once lived in Edinburgh and Obama's forming the belief that I once lived in Edinburgh, is mediated by intervening facts about Obama's mental states.

In general, then, if a thinker forms a belief in normal circumstances (that is, the kind of circumstances that are presupposed by the principles of folk psychology), the proximate psychological explanation of her forming this belief must be some fact about her mental states and events. In fact, it is plausible that this is one of the distinctive features of forming beliefs, in contrast to acquiring mental states of other kinds. Even in the most normal circumstances, one might become hungry, or start feeling a pain, even if the proximate explanation of one's acquiring these mental states is a fact about the state of one's body, and does not involve any fact about one's mental events and states at all; but for the formation of a belief to be explicable in

²⁶ The argument that I give here is a generalization of an argument that I gave elsewhere (Wedgwood 2002).

the normal way, according to the general principles of folk psychology, its proximate explanation must be a fact about the thinker's mental events and states.

So far, I have only argued that the proximate explanation of a thinker's forming a belief, in circumstances that count as normal according to the principles folk psychology, must be some fact about the thinker's mental states and events. I shall now argue more specifically that this proximate explanation must consist in an *internal* fact about the thinker's mental states.

At this point, it becomes important to be clear about what exactly is the *explanandum* in the explanations that we are focusing on. It turns out that there are two conceptions of this *explanandum* that we need to consider. On the first conception, the explanandum is simply the thinker's forming a belief in a particular proposition *p*. On the second conception, the *explanandum* is not identified independently of the whole explanation; rather, the idea is that there is some relation that the proposition *p* stands in to the thinker's situation at the relevant time, such that the *explanans* is the fact that the thinker is in that situation, and that *explanandum* is the fact that the thinker forms a belief in a proposition that has that relation to the situation.

In the remainder of this section, I shall argue for the following two points. First, on the first conception of what the relevant *explanandum* is, the proximate explanation of the thinker's forming a belief in this proposition *p* must be a fact about the thinker's mental events and states that is – in the terms that I introduced in Section 3 – *no less internal* than the *explanandum* itself. Secondly, on the second conception of the relevant *explanandum*, the explanandum and the explanans must both consist of facts about the thinker's mental events and states that are *maximally* internal – that is, facts that have the highest of the three degrees of internality that I enumerated in Section 3.

According to the first point that I shall be arguing for, then, if the *explanandum* is the fact that a thinker forms a certain belief at a certain time, the proximate explanation of this fact must be a fact about the thinker's mental events and states that is no less internal than the *explanandum*. For example, the proximate explanation of the fact that one believes *p* may involve a fact about which propositions one *believes*, but it cannot involve a fact about which propositions one *knows* (since facts about what one knows are “more external” than facts about belief).²⁷

In arguing for this point, I shall assume the correctness of the suggestion that I made at the end of the Section 3, that the relation between a broad mental state and the corresponding internal state is the relation of a *determinate* to a *determinable*. Thus, for example, the broad state of knowing *p* is a determinate of the determinable internal state of believing *p*; and likewise, the state of believing *p* is a determinate of the still more internal determinable state of believing a proposition of the narrow type *T*.

So, consider a case in which the *explanandum* – the fact that we are trying to explain – is the fact that the thinker forms a belief in a given proposition *q*. Now consider two possible

²⁷ This is not to deny that ordinary folk-psychological discourse will often cite the fact that a thinker is in a factive mental state as the proximate explanation of a belief. (“Why did she form a belief in *p*? Well, she saw it with her own eyes!”) In these cases, I suggest, we are only indicating the correct proximate explanation loosely and imprecisely, while simultaneously giving other information that is useful to our audience in the context. For example, by explaining the thinker's forming a belief in *p* on the basis of her seeing with her own eyes that *p* is the case, we are simultaneously indicating the correct proximate explanation of her forming the belief (she formed the belief because she had an experience as of *p*'s being the case), and also indicating a more remote explanation (she formed the belief because *p* was visibly true in her environment).

explanations of this fact. According to the first of these explanations, the thinker forms this belief because she is in the state of *knowing* a certain proposition *p*. According to the second explanation, she forms this belief because she is in the internal state of *believing* this proposition *p*. (We might have focused on a different second explanation, according to which she forms this belief because she is in the internal state that, according to internalism, can be identified with the state of *rationally believing p*. But I shall keep things simple, by focusing on the explanation that appeals to the thinker's believing *p* as opposed to her rationally believing *p*.)

Suppose that each of these two explanations is attempting to identify the *proximate* explanation of the *explanandum* in question; moreover, both explanations assume that the circumstances are normal, in the way that is typically presupposed by folk-psychological explanations. As we have seen in our study of the argument from hallucination, the reason for recognizing that such internal states exist is precisely because normally their immediate short-term causal effects are so similar to those of the corresponding broad states. So, in normal circumstances, this internal state of believing *p* has the same immediate, short-term effects as the broad state of believing *p*. Since these explanations are attempting to identify the proximate explanation of the *explanandum*, only the immediate, short-term effects of the state that is cited as the *explanans* are relevant for the truth of the explanation. Thus, the fact that the thinker is in the internal mental state of believing *p* will be just as close to being causally sufficient for the *explanandum* – forming a belief in *q* – as the fact that she is in the broad state of knowing *p*. In other words, the conditional chance that the thinker will form a belief in *q* is just as high given that she is in the antecedent internal state of believing *p* as the conditional chance that she will form a belief in *q* given that she is in the antecedent broad state of knowing *p*.

However, the second explanation will obviously be *more general* than the first. There could easily be a second case in which the first explanation is not true of the thinker, but the second explanation still is true. That is, in a second case of this sort, the thinker does *not* know *p*, but still believes *p*. As we have seen, the immediate short-term mental effects of believing *p* are extremely similar to the short-term effects of knowing *p*. So, in this second case, the thinker would still form the belief in *q*. So the second explanation – which appeals to the fact that the thinker believes *p*, rather than to the fact that the thinker knows *p* – applies to a significantly wider range of cases.

It is a plausible general principle that other things equal, we should prefer the more general of two explanations that otherwise count as equally good explanations of the same effect, from the same temporal distance. This point is especially plausible if the fact cited as the *explanans* in the more general explanation is a determinable of which the fact cited as the *explanans* in the less general explanation is a determinate. Here is a simple illustration of this point. Suppose that we want to explain why a certain code-protected door opened for the hero. One explanation that we could give would be to say that the door opened because the hero drew an equilateral triangle with each side measuring three inches, using her right index finger. A second explanation that we could give would be to say that the door opened because she drew a triangle. Now suppose that in fact *any* triangle drawn on the code-pad would have succeeded in opening the door. In that case, the second explanation is a better explanation, because it is more general than the first.²⁸

²⁸ I owe this example to Robert Stainton. For some further discussion of this principle about why we should under certain circumstances prefer the more general causal explanations, see Yablo (1992a, 413–23, 1992b, and 1997).

The overall effect of this general principle about explanation is that in any correct explanation there must be a certain sort of *proportionality* between the *explanandum* and the *explanans*. The *explanans* must be sufficient in the circumstances to produce the *explanandum*; but it also must not contain any irrelevant elements that could be stripped away without making it any less sufficient to produce the *explanandum* (and without detracting from the explanation in any other way). What this principle implies for the case at hand is that if in normal circumstances *either* the fact that the thinker knows *p* *or* the fact that the thinker believes *p* is the proximate folk-psychological explanation of the thinker's forming a belief in *q*, it will be the fact that the thinker believes *p*, and not the fact that the thinker knows *p*, that counts as the proximate explanation. In general, in normal circumstances of this kind, the proximate folk-psychological explanation of these cases in which a thinker forms a belief is itself always a fact about the thinker's mental states that is no less internal than the fact of the thinker's forming that belief.

This is not to say that broad states *never* play a role in psychological explanations. As Timothy Williamson has persuasively argued, knowledge does seem to play such a role in the explanation of certain *actions*: Williamson's (2000: 62) compelling example involves a burglar who ransacks the house all night long, because he knows that the house contains a diamond. Here, however, the *explanandum* consists in a thinker's interacting with his environment in a certain way, over an extended period of time. It is only to be expected that the *explanans* – the burglar's knowing that the house contains the diamond – will also consist in the thinker's standing in a certain relation to his environment. This does not show that such broad states will figure as the *proximate* explanation of a comparatively internal fact – like the fact that the thinker forms a belief in *q* at a particular time *t*. As we have seen, in normal circumstances, an internal mental fact of this sort will have a correspondingly internal fact about the thinker's mental states as its proximate folk-psychological explanation.

So far, I have just considered explanations in which the *explanandum* is given in advance as the fact that the thinker forms a belief in a particular proposition *q*. As I noted above, however, there is also a second conception of the kind of explanation that is relevant. On this second conception, the precise *explanandum* is not identified independently of the explanation; rather, the idea is that there is some relation that the proposition *q* stands in to the thinker's situation at the relevant time, such that the *explanans* is the fact that the thinker is in that situation, and that *explanandum* is the fact that the thinker forms a belief in a proposition that has that relation to the situation. The relevant explanation is the best explanation of this kind.

The principle about explanation that was discussed above implies that other things equal, more general explanations are better explanations. So it seems that the best explanation of the thinker's forming the relevant belief will be an explanation that is even more general than any explanation that mentions the particular broad content of that belief. Instead, it will be an explanation that applies, not just to cases in which the thinker's beliefs count as knowledge as well as cases in which they do not, but also to cases in which the contents of the thinker's mental states concern the environment of Earth as well as cases in which the contents of these states concern the environment of Twin Earth. This explanation will be the most general explanation of the phenomenon in question. Clearly, it would have to be an explanation in which the proximate explanation of the thinker's forming the relevant belief is a fact about the thinker's mental states that is internal to the *highest* of the three degrees that I distinguished above – a fact that is present not just in the evil-demon world as well as the normal-perception world, but also in the world in which one is on Earth as well as the world in which one is on Twin Earth.

In the previous section, I argued that rationality must be capable of *directly guiding* your reasoning, and that this implies that the fact that it is rational for you to form a certain belief at a certain time must supervene on facts that are capable of being the proximate folk-psychological explanation of your forming that belief at that time. More precisely, this proposal has two parts. First, the fact that it is rational for you to form a belief in this particular proposition p must supervene on facts that are capable of being the proximate folk-psychological explanation of your forming such a belief in p . Secondly, when it is rational for you to form a belief in a certain proposition, that is because the proposition stands in a certain relation to your situation, and the fact that your situation has that relation to a proposition must be capable of being the proximate folk-psychological explanation of your forming a belief in a proposition that stands in this relation to your situation.

We have now seen that the only facts capable of being the proximate folk-psychological explanation of a thinker's forming a certain belief at a certain time are internal facts about the mental events states that are present in the thinker's mind at that time. Exactly *how* internal these facts have to be depends on whether the *explanandum* consists of the thinker's forming a belief in the particular proposition in question, or just of the thinker's forming a belief in some proposition or other that has the relevant relation to the thinker's situation.

Two conclusions follow from what has been argued so far. First, the fact that in your situation it is rational for you to form a belief in p must be a fact about your mental events and states that is internal to the intermediate degree (that is, it must be a fact that does not involve any factive mental states like knowledge, though it may be a fact that can vary between Earth and Twin Earth). So the property that your situation has, of being a situation in which it is rational for you to form a belief in p , supervenes on facts that are internal to this intermediate degree.

Secondly, whenever it is rational for you to form a belief in a certain proposition, there is some fact to the effect that it is rational for you in this situation to form a belief in a proposition that has a certain relation to your situation, such that every such fact is a fact about your mental events and states that is internal to the highest degree. Call such facts "*R*-facts". Take any pair consisting of a situation that you can be in and a belief that it is rational for you to form in this situation. Then there will be an *R*-fact that holds in that situation, and since this *R*-fact is a maximally internal fact, every situation-belief pair that resembles the pair in question in all maximally internal respects will also involve a situation in which this *R*-fact holds. In other words, the property that this pair has, of being a situation-belief pair such that it is rational in the situation to form the belief, supervenes on maximally internal facts of this sort.

This then is the explanation of why internalism is true – that is, of why it is that the facts about what is rational and about what is not rational supervene on internal facts about the mental events and states that are present in the thinker's mind at the relevant time. The details of this account have been a little tricky, but the basic picture is clear. The facts about what is rational and what is not rational must be capable of directly guiding a rational thinker. But in the relevant sense, only internal facts about the mental states and events that are present in the thinker's mind at the relevant time can directly guide the thinker. It follows that the facts about what is rational and what is not rational must be internal facts of this sort. In this way, an explanation can be given of why an internalist conception of rationality is true.

7. Conclusion: Rationality and the past

The explanation of internalism given above helps to answer yet another tricky question that arises in the discussion of rationality. On some views – such as the view that has recently been defended by Brian Hedden – whether your mental states are rational at a particular time is determined purely by what is going on in your mind at that very time; as Hedden (forthcoming, 1) puts it, “locus of rationality, so to speak, is the time-slice rather than the temporally extended agent”. On other views – such as the view that was defended by Goldman (1999: 280f.) – long-forgotten facts about how one’s belief was originally acquired years ago may still be part of what determines whether the belief is justified now; and what is present within the mind at the present time may not be sufficient to determine whether the belief is justified or not.

The explanation of internalism given above seems to suggest that an intermediate position, between the extreme positions of Hedden and Goldman, is more plausible here. The rationality of your mental states and events supervenes on facts that are, in the sense that I explained, capable of directly guiding your thinking. It seems that your thinking can be directly guided by facts about the mental states and events that were present in your mind in the *immediate* past (although it is clearly an empirical question how far back into the past your mental states can be while still being capable of guiding you directly in the relevant way); but facts about what was present in your mind in the more remote past are capable of guiding your thinking only *indirectly*, by means of guiding some of your thinking during the intervening period of time.

So, it seems plausible to me that the rationality of your mental states supervenes on what is present in your mind both now and in this short period of the immediate past – but that it is independent of what was present in your mind in the more remote past. For example, consider a being who is brought into existence as a molecule-for-molecule duplicate of you, in the state that you were in on the morning of your eighteenth birthday. It is disputed whether or not this being would have mental states from the very first moment of its existence, but we do not need to take a stance on this question. Suppose that there is a first moment at which this being has mental states, and that from that point on, this being’s mental states and events are exactly like yours. According to the explanation of rationality that I have given here, at the very first moment when this being has mental states, these mental states may well be arational – neither rational nor irrational – since without any facts about the thinker’s prior mental states, nothing can determine whether or not these mental states are rational. However, after this initial point, exactly the same mental states are rational in your case as in the case of your more recently-created duplicate.

In this way, my explanation of internalism about rationality seems to illuminate some of these difficult questions, in an illuminating and intrinsically plausible way. This strengthens our reasons for thinking that this explanation is along the right lines – and consequently that the intuitive argument for internalism that we started with is sound.

References

- Alston, William (1989). *Epistemic Justification* (Ithaca, NY: Cornell University Press).
 Antony, Louise M. (1995). “I’m a mother, I worry”, *Content* 6: 160–166.
 Ayer, A. J. (1973). *The Central Questions of Philosophy* (London: Weidenfeld).
 Cohen, Stewart (1984). “Justification and truth”, *Philosophical Studies* 46 (3): 279–95.

- Corrado, Michael L. ed. (1994). *Justification and Excuse in the Criminal Law* (New York: Garland).
- Comesaña, Juan (2002). “The diagonal and the demon”, *Philosophical Studies* 110 (3): 249–266.
- Dancy, Jonathan (1995). “Arguments from illusion”, *Philosophical Quarterly* 45 (181):421–438.
- Danto, Arthur (1968). “Basic Actions”, in Alan White, ed., *The Philosophy of Action* (Oxford: Oxford University Press): 43–58.
- Goldman, Alvin (1979). “What is Justified Belief?”, in G. S. Pappas, ed., *Justification and Knowledge* (Dordrecht: Reidel, 1979).
- (1986). *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
- (1999). “Internalism Exposed”, *Journal of Philosophy* 96: 271–93.
- Hedden, Brian (forthcoming). “Time-Slice Rationality”, *Mind*.
- Hinton, J. M. (1973). *Experiences* (Oxford: Clarendon Press).
- Johnston, Mark (2004). “The Obscure Object of Hallucination”, *Philosophical Studies* 120: 113–183.
- Loar, Brian (2003). “Phenomenal Intentionality as the Basis of Mental Content,” in Martin Hahn and Bjørn Ramberg, ed., *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge, MA: MIT Press).
- McDowell, John (1994). *Mind and World* (Cambridge, MA: Harvard University Press).
- Millar, Alan (1996). “The Idea of Experience”, *Proceedings of the Aristotelian Society* 96: 75–90.
- Plantinga, Alvin (1993). *Warrant: The Current Debate* (New York: Oxford University Press).
- Pollock, J. L., and Cruz, Joseph (1999). *Contemporary Theories of Knowledge*, revised edition (Totowa, NJ: Rowman and Littlefield).
- Pryor, James (2001). “Highlights of Recent Epistemology”, *British Journal for the Philosophy of Science* 52: 95–124.
- Putnam, Hilary (1975). “The Meaning of ‘Meaning’”, reprinted in Putnam, *Mind, Language, and Reality* (Cambridge: Cambridge University Press).
- Sosa, Ernest (1994). “Philosophical Scepticism and Epistemic Circularity”, *Aristotelian Society Supplementary Volume* 68: 263–290.
- Snowdon, P. F. (1981). “Perception, Vision and Causation”, *Proceedings of the Aristotelian Society* 81: 175–192.
- Wedgwood, Ralph (2002). “Internalism Explained”, *Philosophy and Phenomenological Research* 65: 349–369.
- (2006). “The Normative Force of Reasoning”, *Noûs* 40, no. 4 (December), 660–686
- (2014). “Rationality as a Virtue”, *Analytic Philosophy* 55, no. 4 (December): 319–338.
- Williamson, Timothy (2000). *Knowledge and its Limits* (Oxford: Clarendon Press).
- Yablo, Stephen (1992a). “Cause and Essence”, *Synthese* 93: 403–49.
- (1992b). “Mental Causation”, *Philosophical Review* 101: 245–80.
- (1997). “Wide Causation”, *Philosophical Perspectives* 11: 251–81.