

Justifications, Excuses, and Sceptical Scenarios

(draft of paper to appear in Julien Dutant and Fabian Dorsch, eds., *The New Evil Demon*,
Oxford: Oxford University Press)

Timothy Williamson

Here are two assumptions about justified belief widespread in contemporary epistemology and often used in explaining what justification is. First, a belief is justified if and only if it meets the key normative standard for the epistemic assessment of beliefs. Second, the beliefs of a subject in a good case and the corresponding bad case are equally justified. For instance, in the good case, I can see my hands, I know that I have hands, and I have a justified belief that I have hands. In the bad case, I am a brain in a vat but I receive the same local inputs as in the good case: since I lack hands, I do not know that I have hands, but I still believe that I have hands, and by the second assumption my belief is equally justified. Consequently, in the bad case, I have justified false beliefs. One can state a consequence of this package of views without using the word ‘justified’: in some cases, a false belief meets the key normative standard for the epistemic assessment of beliefs. Whether that consequence is true or false, it is too important to be smuggled into an explanation of the phrase ‘justified belief’.

Issues about the meaning of the phrase ‘justified belief’ also arise concerning recent methodological challenges to the use of Gettier cases in epistemology. Those challenges have mainly focussed on the status of the judgment that the true belief in question is not knowledge. Much less discussed has been the status of the judgment that the belief is justified. For better or worse, ‘She has a justified belief’ sounds more theoretical, less ordinary, than ‘She knows’. I am not the only epistemologist to have had the experience, in explaining a traditional Gettier case to non-philosophers, of finding that while they happily accept that the belief is not knowledge, they deny that it is justified. Sometimes they protest that the subject is not justified in relying on a false assumption. Although one can browbeat them into accepting that the belief is justified by getting them to see the question only from the subject’s point of view, the obvious danger is that all they may then be noticing is that the belief *seems* justified from the subject’s point of view, not that it really is justified. We should not assume without argument that there is no distinction between appearance and reality for justification. After all, perhaps a sociologist of science can browbeat people into accepting that pre-Copernicans *knew* that the earth is the centre of the universe, by getting them to see the question only from the pre-Copernicans’ point of view.

I do not suggest that *most* lay people deny justification in Gettier cases. In an experimental investigation, Jennifer Nagel, Valerie San Juan, and Raymond Mar (2013) found that subjects tended to rate the justification of the crucial belief in Gettier cases quite high, and that the tendency was correlated positively with a measure of empathy.¹ If the assessment of justification depends on whether one views the case in a first-personal or third-personal way, we need to understand what sort of contrast is in play. When non-

philosophers classify the Gettier belief as justified, what do they take that to imply? Does it depend on what other classifications are on offer?

Some epistemologists may object that 'justified' is used in a *special sense* in epistemology, and that laypeople must be taught this special sense, perhaps by means of examples, before they can engage with the Gettier problem as usually formulated. Indeed, epistemologists usually explain that they are speaking of *epistemic* rather than *pragmatic* justification: if believing that you will win a race raises your chance of winning from 10% to 30%, you may thereby be pragmatically justified in so believing, but you are not epistemically justified. 'Epistemic justification' is manifestly technical terminology: we should be correspondingly suspicious of claims to make pre-theoretic judgments about its application. Moreover, we might then query the interest of the Gettier problem, if it merely shows that one cooked-up sense of 'justified' does not serve the purpose for which it was cooked up. Although merely pragmatic justification does not seem to be at issue in standard Gettier cases, 'epistemic justification' itself can be understood in a variety of ways.²

The Gettier problem is most significant when formulated with the word 'justified' understood in something close to its usual normative sense. In that case, denials of justification from outside the epistemology club are at least *relevant*, even if mistaken. To clarify the issues, we must identify the relevant norm(s): in this case, epistemic norm(s), by the setup of the problem. To identify the norm(s), we must be sensitive to standard normative distinctions, such as that between justification and blamelessness. In killing a child who suddenly runs out into the road, a driver may be blameless without being justified. Thus we may ask whether a brain in a vat is justified, or merely blameless, in

believing (falsely) that it has hands.³ Similarly, is someone who relies on a stopped clock justified, or merely blameless, in believing (truly but unknowledgeably) that it's 3 p.m.? Do they have a justification or just an excuse? Are they complying with the relevant norm, or violating it in mitigating circumstances? That depends, of course, on what the norm is. Epistemologists routinely use normative terms such as 'justified', 'rational', and 'reasonable', but too often neglect vitally relevant, though subtle, normative distinctions.

This essay steps back from the epistemological issues to make some of those general normative distinctions, then returns with them to epistemology.

2. As is usual in epistemology and the philosophy of language, I will apply the word 'norm' liberally, to anything that yields some sort of 'ought' or 'should'. In particular, I will not restrict it to *moral* norms. It would be inappropriate to assume that the distinction between justified and unjustified beliefs is a moral distinction. One epistemically ought not to have unjustified beliefs, even when one morally ought to have those very beliefs, because in the circumstances having them is the only way of convincing the secret police that one has them, which is in turn the only way of saving one's family. Similarly, the distinction between legal and illegal moves in a game is a normative distinction. With respect to the rules of the game, one ought not to make illegal moves, even when one morally ought to make those very moves, because one is playing a concentration commander at cards for the life of a fellow inmate, and cheating is the only way of saving her life.⁴

I will follow the usual practice of sometimes using the word ‘norm’ for *candidate* norms, putative norms that may or may not be genuine norms (may or may not yield some sort of genuine ‘ought’ or ‘should’), when no confusion is likely to arise.

A norm divides cases into those in which it is *complied with* and those in which it is not (it is *violated*). Thus a norm has a compliance condition. For illustrative purposes, it will serve clarity to start with some norms whose specific compliance conditions are explicit in particular circumstances. The norm of promise-keeping is an example. Here ‘keeping a promise’ is *not* read as entailing ‘intentionally keeping a promise’, just as ‘complying with a norm’ does not entail ‘intentionally complying with the norm’ and ‘obeying a law’ does not entail ‘intentionally obeying the law’, on the relevant readings. On such readings, there is often a conversational implicature that the ‘doing’ was intentional, even though there is no entailment to that effect. For the sake of clarity, I hereby cancel any such implicature for present purposes.

Promises make their specific compliance conditions explicit. Of course, the *general* compliance condition for the norm of promise-keeping is just that one keeps one’s promises. But when one promises, *what* one promises is a compliance condition specific to that promise. For instance, if I promise to be in London on Monday, then I comply with the norm of promise-keeping with respect to that promise if and only if I am in London on Monday, for whatever reason. If I am not in London on Monday, for whatever reason, then I violate the norm of promise-keeping, I fail to comply. Even if I intend to keep my promise, try my hardest to keep it, perhaps believe on strong evidence that I *am* keeping it, but am not in London on Monday, then I have not kept my promise, I have broken it. For I did not do what I promised to do. What I promised was not to do

my best to be in London on Monday, but simply to *be in London on Monday*. I could have promised to do my best to be in London on Monday, but that would have been a different promise, with different compliance conditions. Again, if I am elsewhere on Monday for overriding moral reasons of manifestly much greater importance than my promise, then I have still broken my promise, and so violated the norm of promise-keeping (unless the person to whom I made the promise released me from it). Conversely, if I forget my promise or intend to break it, but am in London on Monday anyway for some other reason, then I have kept my promise, and so far complied with the norm of promise-keeping, perhaps unintentionally. Promising to be in London is quite different from promising to do my best to be in London, and from promising to be in London unless something more important turns up. In general, if I promise to ϕ , then I keep my promise if and only if I ϕ , and I comply with the norm of promise-keeping with respect to that promise if and only if I do keep it.

Of course, if I promise you that I will be in London on Monday, make no effort to be in London on Monday, but end up there then merely by accident, you may still complain that I made no effort to keep my promise. But the primary obligation to you I undertook in promising was just to be in London on Monday. You are complaining that I made no effort to fulfil my primary obligation. Any obligation to make an effort to fulfil one's primary obligation is a secondary obligation, derivative from the first. After all, I *promised to be* in London on Monday; I did not *promise to make an effort* to be in London on Monday. Lumping the two obligations together under some inflated conception of 'promise-keeping' would merely obscure the structural dependence of one on the other.

Another norm with explicit compliance conditions in particular circumstances is the norm of obeying the law. In circumstances governed by a given code of law, that code makes explicit the specific compliance conditions of the norm. Obeying the law is quite different from doing one's best to obey the law, and from obeying the law unless something more important turns up. My intentions are relevant to whether I obey a law only if its specific content makes them relevant. They are relevant to whether I obeyed a law not to be intentionally in the Holy of Holies, but irrelevant to whether I obeyed a law simply not to be in the Holy of Holies. In general, if the law is that one must ϕ , then I obey that law if and only if I ϕ , and I comply with the norm of obeying the law with respect to that law if and only if I do obey it.

Let N be a norm for agents in situations. In a given situation, an agent may comply with N, or fail to do so. A given norm typically generates various derivative norms. In particular, there is the secondary norm DN of having a *general disposition* to comply with N, of being the sort of person who complies with N. Whether an agent A complies with DN in a situation S depends not just on whether A complies with N in S but also on whether A complies with N in possible situations beyond S. For example, if L is the norm of obeying the law, then DL is the norm of being law-abiding, of being the sort of person who obeys the law. If P is the norm of keeping one's promises, then DP is the norm of being the sort of person who keeps their promises (a 'promise-keeper' for short). DN is the natural norm derivative from N on the subject's dispositions.

As the examples make clear, DN need not inherit the full normative status of N. If one violates L, one *ipso facto* breaks the law. If one violates DL, for accidental reasons one might still never break the law, at least in a non-totalitarian state where lacking a

law-abiding disposition is not itself illegal. Thus L in effect has the force of law, while DL has not. Similarly, if one violates P, one *ipso facto* breaks a promise. If one violates DP, for accidental reasons one might still never break a promise, unless for some reason one promises to have a promise-keeping disposition. Thus P in effect has the quasi-contractual status of a promise, while DP has not. Nevertheless, some of the normative status of N typically rubs off on DN. Usually, if you ought to comply with N, then you ought to comply with DN too, even though the underpinning of the latter ‘ought’ is derivative from, and more complicated than, the underpinning of the former ‘ought’. We can therefore call DN a norm too.

On the intended reading of the phrase ‘the sort of person who complies with N’, ‘complies with N’ is to be understood as generic. The sort of person who complies with N need not be the sort of person who complies with N in all possible situations; or even in all actual situations. Interfering circumstances sometimes prevent a disposition from manifesting itself. The sort of person who complies with N is the sort of person who *normally* complies with N, but may fail to do so in abnormal cases — and perhaps occasionally in normal ones too, because even the best of us can undergo a blip.

More formally, complying with N in a situation S is neither necessary nor sufficient for complying with DN in S. For instance, I promised to give you the money, but now I want to give it to Mary instead. For some reason, you are perfectly disguised as Mary and Mary is perfectly disguised as you. If I am a promise-keeper, then I shall give the money to Mary, believing her to be you; I thereby violate P, while complying with DP. Conversely, if I am the sort of person who breaks their promises whenever it suits

them (a ‘promise-breaker’ for short), then I shall give the money to you, believing you to be Mary; I thereby comply with P, while violating DP.

There is also a tertiary norm ODN of doing what someone who complied with DN would do in the situation at issue (‘O’ for ‘occurrent’). For example, I comply with ODP in a situation S if and only if I do in S what a promise-keeper would do in S. The status of ODN is derivative from that of DN, and thus doubly derivative from that of N.

We may assume, at least for the sake of simplicity, that complying with DN entails complying with ODN. For if you comply with DN, then what you would do in the situation S is what *someone* who complies with DN would do in S, even if it is not what *everyone* who complies with DN would do in S.⁵ That leaves six combinations of compliance and violation with respect to N, DN, and ODN, all of which are possible in principle. To reinforce the distinctions, they are illustrated here with respect to promise-keeping.

(i) Complying with P, DP, and ODP: I am a promise-keeper; in a very ordinary situation I make a promise and keep it.

(ii) Complying with P and ODP while violating DP: I am a promise-breaker, but in a very ordinary situation I make a promise and keep it because it happens to suit me to do so.

(iii) Complying with P while violating DP and ODP: This is the case above where I am a promise-breaker but you and Mary disguise yourselves as each other.

(iv) Violating P while complying with DP and ODP: This is the case above where I am a promise-keeper but you and Mary disguise yourselves as each other.

(v) Violating P and DP while complying with ODP: In a variant of the case above, I am a promise-breaker, decide to give you the money after all because it suits me to do so, but as a result give it to Mary instead.

(vi) Violating P, DP, and ODP: I am a promise-breaker; in a very ordinary situation I make a promise and break it.

Cases (ii), (iii), and (iv) show that P is not equivalent to DP; P is neither necessary nor sufficient for DP. Cases (iii), (iv), and (v) show that P is not equivalent to ODP; P is neither necessary nor sufficient for ODP. Cases (ii) and (v) show that DP is not equivalent to ODP; DP is not necessary for ODP.

Despite such differences between the three norms, if we endorse a primary norm N, we shall probably be sympathetic to the corresponding secondary norm DN and the tertiary norm ODN too. If it is bad not to keep a promise, then there is also something bad about not being generally disposed to keep one's promises, and something bad about not doing what someone so disposed would do. But that does not mean that the three norms have equal status. Typically, any normative significance that DN possesses is merely derivative from that of N, and any normative significance that ODN possesses is merely derivative from that of DN and thence from that of N, as in the case of promise-keeping.

The point can be put in terms of the distinction between justifications and excuses. The distinction, whose fruitfulness for philosophy Austin memorably emphasized, has been much contested in jurisprudence. For present purposes, it can be given a quite straightforward form.⁶ We consider justification with respect to a given norm; although *moral* justification may be an all-things-considered matter, the case of epistemic justification makes it inappropriate to treat justification in general as all-things-

considered. Then, with respect to the given norm, one is justified if and only if one complies with that norm. Excuses arise only when one violates the norm. They are pleas in mitigation of the offence. With a good enough excuse, one is blameless but not justified. In cases (i), (ii), and (iii) above, one complies with N and is justified with respect to N. In cases (iv), (v), and (vi), one violates N and is not justified with respect to N. When one violates N but complies with ODN (cases (iv) and (v)), one's compliance with ODN is an excuse but not a justification with respect to N. Violating N is never justified with respect to N, although it may be justified with respect to some other norm.

In this framework, justification is neither necessary nor sufficient for *blamelessness*. Justification is not necessary for blamelessness because in case (iv) the promise-keeper blamelessly breaks a promise, and so is not justified. Justification is not sufficient for blamelessness because in cases (ii) and (iii) I am justified with respect to P, since I keep my promise, even though I am blameworthy because I violate ODP and, in case (iii), DP too. For the sake of clarity, I will restrict the term 'justification' to compliance with the relevant *primary* norm (whether epistemologists use the term so strictly is another matter). Thus, for present purposes, given that the normative status of DN and ODN is merely derivative from that of N, there is no such thing as justification with respect to DN or ODN. If I promised to give you the money but gave it to Mary instead, that I am a promise-keeper but took Mary for you, and so did what a promise-keeper would do in the circumstances, is an excuse rather than a justification for not keeping my promise, for not giving you the money. Being a promise-keeper and doing what a promise-keeper would do in the circumstances have a merely derivative normative status.⁷

Compliance with ODN is usually a much better excuse for violating N if one also complies with DN. That one was doing what a person of good character would do in the circumstances is a more impressive excuse if one is indeed of good character than if one is not. The excuse that one was doing what a promise-keeper would do sounds better in the mouth of a promise-keeper (case (iv)) than in the mouth of a promise-breaker (case (v)).

3. We can apply the apparatus just developed to epistemology. We may start with a wide class of (putative) norms for belief which I will call ‘truth-related’, for want of a better term. Truth-related norms for belief constrain one’s current beliefs, rather than one’s dispositions to form or maintain beliefs (they are not of the form DN). Truth-related norms for belief include norms such as these: that one’s beliefs should be true; that one’s beliefs should constitute knowledge; that one’s beliefs should be probable on one’s evidence; that one’s beliefs should be consistent.⁸ These norms differ amongst themselves in obvious ways. For example, the brain in the vat as usually envisaged violates the truth and knowledge norms but complies with the probability and consistency norms. Nevertheless, despite their differences, each of those norms imposes a truth-related standard on the content of a belief, a norm that is somehow objective. Trying one’s hardest to comply with such a norm is neither necessary nor sufficient for complying with it. As a minimal requirement, we may stipulate that a norm for belief counts as truth-related only if it excludes belief in an explicit contradiction. All four norms just mentioned meet that requirement: a contradiction cannot be true, and so cannot be known, it has probability zero, and it is inconsistent.⁹

Let J be a truth-related norm of belief. Then DJ is the norm of being the sort of person who complies with J, and ODJ is the norm of doing in the given situation what the sort of person who complies with J would do. On general grounds, we may expect the primary norm J, the secondary norm DJ, and the tertiary norm ODJ to be as usual three mutually non-equivalent norms. Nevertheless, those sympathetic to J as an appropriate norm for belief will probably feel some sympathy for DJ and ODJ too as norms for belief, even if in the end they give primacy to J. They may be tempted to call beliefs ‘justified’ because they comply with DJ and ODJ, although only compliance with a primary norm confers justification in the sense used here.

A Pyrrhonist sceptic may hope to comply vacuously with all three norms by having a general disposition never to believe anything. If one has no beliefs, then *a fortiori* one has no untrue beliefs, no beliefs that fail to constitute knowledge, no beliefs that are improbable on one’s evidence, no inconsistent beliefs, and so on. The Pyrrhonist, if such a person is possible, complies with all three norms even in the sceptical scenario. Since complying with ODJ is doing what the sort of person who complies with J would do, compliance with ODJ depends on whether that sort of person is a Pyrrhonist. In what follows, that sort of person will be assumed *not* to be a Pyrrhonist, but instead to be a more normal person, who like the rest of us would have false beliefs in a sceptical scenario. Non-sceptics may find little to admire in the Pyrrhonist’s self-imposed ignorance, especially when that ignorance concerns the needs of others. There may be positive norms for knowledge, such as a norm enjoining knowledge-gathering in various circumstances, and so positive as well as negative norms for beliefs.¹⁰

Even if the victims of sceptical scenarios violate J, we may still expect them to comply with both DJ, and so with ODJ. For example, although most of their beliefs are false and so not knowledge, we may assume that they are still the sort of people whose beliefs are *normally* knowledge and so *normally* true, at least when they concern straightforward features of the local environment, and therefore that in the sceptical scenario they do what such people would do. For even the general dispositions of the brain in the vat remain truth-tracking and knowledge-producing ones, we may suppose, although the unfavourable circumstances block those manifestations. In the relevant sense of ‘normal’, which is not a purely statistical one, sceptical scenarios are *abnormal* cases. Thus, even if the victims of sceptical scenarios violate a primary truth-related norm of justification, we are still likely to feel some residual sympathy for the claim that their beliefs are justified, because they comply with the associated secondary and tertiary norms of justification, even though strictly speaking compliance with a derivative norm provides only an excuse, not a justification.¹¹

The justified true belief analysis of knowledge allows for justified untrue belief, since otherwise ‘true’ would be redundant. Most of its critics, from Gettier onwards, have allowed for justified false belief.¹² Indeed, many analytic epistemologists say that the victims of sceptical scenarios are every bit as justified in their beliefs as they would be in the corresponding good cases, where those beliefs are knowledge.¹³ This supposed equality in justification between the good and bad cases is sometimes used to *explain* what justification is, although our present interest concerns justification independently introduced as some appropriate norm for belief, not definitionally stipulated to be equal between those cases. But the idea seems to be that making us look for features common

to the good and bad cases will draw our attention to a central norm for belief that is insensitive to the difference. In present terminology, it is treated as a primary norm, for someone who treated it as merely secondary or tertiary would hardly be as resistant to describing compliance with it as providing only an excuse, rather than a justification, as many such epistemologists are.

On the orthodox view, the victims of classic sceptical scenarios comply non-Pyrrhonistically with all three of J, DJ, and ODJ even in those scenarios, where J is justification in the intended normative sense. One might then wonder whether J and ODJ coincide after all, so that one has a justified belief if and only if one believes what the sort of person whose beliefs are normally justified would believe?¹⁴ In section 2, we saw that N and ODN differ in their compliance conditions for some norms N, perhaps even for most, but the evidence there did not establish the conclusion that N and ODN differ in their compliance conditions for *all* norms N. A trivial counterexample is that if N is a norm with which it is impossible to comply, then so is ODN, so in that case N and ODN coincide. Less trivially, it can be argued that for many norms N, ODN is itself a fixed point of OD: in other words, $ODODN = ODN$.¹⁵ Thus if justification is itself a norm J of believing as someone disposed to comply with some norm N would believe, then ODJ is equivalent to J.

However, I will now argue that even J and ODJ differ in their compliance conditions. By hypothesis, J is a truth-related norm for belief, and so excludes belief in an explicit contradiction. By contrast, ODJ does *not* exclude belief in a contradiction, and so is not a truth-related norm for belief. To see that, consider this possibility. A special device emits waves of some sort with a selective scrambling effect on brains. The waves

inflict no permanent damage, and do not even change what programme it would be natural to describe the brain as running, but they occasionally alter the contents of unconscious short-term working memory, so that some computations produce incorrect results. Thus a normal subject may confidently announce that $17 + 29 = 33$. Similarly, consider Connie, a normal subject and competent mathematician who sincerely and confidently announces that 179 is and is not prime, because a scrambled piece of reasoning yields that conclusion, and a scrambled application of a contradiction-detector failed to sound the alarm in retrospect. Do not imagine Connie as someone who has gone temporarily mad; rather, think of her on the model of someone doing a long calculation on paper, who fails to notice that a prankster is occasionally rubbing out figures and replacing them by other ones. Connie's attitude to the proposition that 179 is and is not prime even amounts to belief: she acts on it, for example when her career is riding on it in a mathematics test. We may assume that these effects are deterministic: the same fallacy occurs whenever the scrambler is used in the same situation (by the relevant standard: say, a competent calculator embarking on that very calculation).

Connie believes that 179 is and is not prime.¹⁶ In doing so, she still complies with the norm DJ, and *a fortiori* with the norm ODJ. For, we may assume, she complies with DJ in normal circumstances, when the scrambler is switched off, and switching it on does not change her general dispositions. Its effect is to interfere temporarily with the operation of her general dispositions, not to destroy them, just as a fragile vase remains fragile even when it has been wrapped in protective material.¹⁷ Nothing more than stopping the interference (switching off the scrambler, removing the wrapping) is needed to enable the disposition to manifest again, unlike cases where interference temporarily

makes something lose a disposition. Since whether one complies with DJ depends only on one's general dispositions, Connie still complies with DJ when the scrambler is switched on. Moreover, in coming to believe that 179 is and is not prime, she does what the sort of person who complies with DJ would do in her circumstances, which include the scrambler interfering in its predetermined characteristic way. Thus, in believing that 179 is and is not prime, she also complies with ODJ. But of course, in believing that 179 is and is not prime, she does *not* comply with the original norm J itself, which by hypothesis excludes belief in an explicit contradiction, because it is a truth-related norm for belief. Thus the norms J and ODJ differ in their compliance conditions, as promised.

The underlying strategy of the argument is not completely reliant on the possibility of belief in a contradiction, or indeed of belief in an impossibility. We could tweak the example into one where the scrambler causes the subject's reasoning to reach, quite fallaciously, a contingently false conclusion which her evidence tells strongly against — say, that most non-Westerners are stupid, or that this candidate deserves the job more than that one does — but which she thereby comes to believe without further reflection. In having that belief, she complies with the norms DJ and ODJ, for the same reasons as before. For, we may assume, she is still the sort of person who complies with the norm J, and in forming the belief that most non-Westerners are stupid she is coming to believe what that sort of person would come to believe in her circumstances, which include the operation of the scrambler. But she does not comply with the original norm J, if that norm is plausibly sufficient for justification in any serious sense. Although her belief that most non-Westerners are stupid does at least satisfy the minimal norm of

consistency, that norm is too weak to be a plausible standard for justification. She is not justified in believing that most non-Westerners are stupid.

The upshot is this. For any plausible norm of justification *J*, there is an associated but non-equivalent tertiary norm *ODJ* of believing what the sort of person who normally complies with *J* would believe in the circumstances. Irrespective of whether normal victims of sceptical scenarios comply with *J*, they comply with *ODJ*. Moreover, anyone sympathetic with the original norm *J* is likely to feel some consequential sympathy for the tertiary norm *ODJ*. The next step is to apply this result to the assessment of the claim that the subject's beliefs are equally justified in the good case (normal circumstances) and the bad case (a sceptical scenario).

4. The term 'justification', as used in epistemology, is supposed to express a centrally important norm for belief. It is used here under the supposition that it does so.

A strongly externalist norm for belief is that one should believe *p* only if one knows *p*.¹⁸ On this view, it is wrong to believe *p* when one does not know *p*. If that is the standard for justified belief, then the brain in the vat's belief that it has hands is unjustified: it does not constitute knowledge, because it is false. What the brain in the vat has is a cast-iron excuse for the belief, not a justification: it is of good cognitive character, has the right general cognitive dispositions, and thereby arrived at its belief, with no accessible reason to doubt that in doing so it complied with all the relevant epistemic norms, no accessible reason to suspect that the belief was unjustified. The belief is therefore blameless, but still unjustified.

As already noted, a commoner view amongst analytic epistemologists is that the brain in the vat's belief that it has hands is every bit as justified as my belief that I have hands. If the only normative options on offer are being justified and being unjustified, one can see why kind-hearted epistemologists would take pity on the brain in the vat and assign its beliefs to the good category, since after all it is doing its best to comply with the norm for belief. However, once the category of blameless but unjustified belief is available, more than charity is needed to make the case for classifying the brain in the vat's beliefs as justified.

Some arguments for the conclusion that the brain's beliefs are justified rely on assumptions that impress only internalists, such as the premise that justification supervenes on internal states of the brain, or on experience, internally individuated. When used to argue against externalism, such premises merely beg the question. I will not discuss them here. Of more interest and wider appeal is a line of thought like this:

(*) In believing falsely that it has hands, the brain in the vat is doing better than mere blamelessness. It is following the very cognitive instincts it ought to have, doing what a well-designed brain should do. If instead it believed truly that it lacked hands, it would be following much worse cognitive instincts, since it has no evidence that it lacks hands. Even if it just suspended belief, and became agnostic as to whether it had hands, it would be following a Pyrrhonian sceptical instinct that no well-designed brain should have, since in normal circumstances it involves self-imposing ignorance. In believing that it has hands, the brain is doing well. Indeed, it is following exactly the same cognitive instincts as it follows in forming

the same belief in the good case. Therefore the belief is justified to exactly the same degree in the good and bad cases.

That line of thought contains much with which an externalist can agree. Nevertheless, the norms to which it appeals are pretty clearly of the general form of DJ and ODJ: the victim of the sceptical scenario has a general disposition to comply with the relevant norm for belief, and its beliefs in the vat manifest that disposition. But, as we saw in the last section, such disposition-based norms are not truth-related norms for belief, and are not plausibly sufficient for justification. For in the brain scrambler case one complies with such norms in believing an overt contradiction, or more generally in believing something for which one has no evidence, on the basis of fallacious reasoning. Such beliefs are not justified, although they may be excusable and even blameless. The brain scrambler case is just one more sceptical scenario, in which the data for computation are interfered with at a slightly different point. Thus the supposed norm to which the line of thought (*) appeals is not a norm for justified belief in any relevant sense.

Even if the standard of justification for belief is in fact knowledge (K), as on the strongly externalist account above, the brain in the vat still complies with the corresponding secondary norm DK and tertiary norm ODK. It is doing what the sort of subject whose beliefs tend to constitute knowledge would do in the circumstances, because it is that sort of subject. The externalist can predict that the line of thought (*) against the strongly externalist account will be tempting, even though it is fallacious. That the line is tempting is no evidence against the strongly externalist account.

The point can be put in terms of the distinction between justifications and excuses. The brain in the vat meets the good excusing condition identified in section 1. Although it violates K, it complies with DK as well as ODK (case (iv)). It is the sort of subject whose beliefs tend to constitute knowledge (on the usual assumption that it is the same subject as the one before the brain was extracted), and it is doing what that sort of subject would do in the circumstances. That is an excellent excuse for violating the norm K, but it is no justification if K is the primary norm.¹⁹

Similar distinctions apply to norms of *rationality*. In a recent paper, Stewart Cohen and Juan Comesaña identify justified belief with rational belief and treat it as obvious that rationality can require one to believe a falsehood. They give an example:

Suppose you notice what appears to be a red table staring you in the face, you have no evidence of deception, everyone else around you says they see a red table, yet you fail to believe there is a red table before you. In our view you are paradigmatically irrational, even if unbeknown to you, you do not see a red table.²⁰

Clearly, if you are the sort of person who is generally disposed to think rationally and non-sceptically, and you exercise those dispositions in this case, then you will believe that there is a red table before you. In other words, you comply with the relevant norm of type ODN. Moreover, if you do not believe that there is a table before you, then you violate the ODN norm, so in the circumstances that norm requires you to believe falsely that there is a red table before you.

To regard rational beliefs as those a rational agent (perhaps an ideally rational agent) would have in the circumstances is to treat rational belief on something like the ODN model. Such a dispositional conception of rationality may be one of the forces pushing many philosophers towards an internalist conception of rationality, for dispositions tend quite generally to be envisaged as internal or intrinsic to the things so disposed. Bubble-wrapping the vase leaves its disposition to break intact. The subject is usually conceived as having the same cognitive dispositions in matching good and bad cases. However, the tendency is not inexorable. Some dispositions are extrinsic. Castles became more *vulnerable* when gunpowder was introduced.

But not all Cohen and Comesaña's claims about rationality fit the ODN model. For instance, they say 'Rationality requires one to conform one's beliefs to one's evidence'.²¹ Complying with the ODN model here would be doing what the sort of person who usually conforms their beliefs to the evidence would do: but if in some cases that sort of person would misjudge what conforming their beliefs to the evidence involves in the circumstances, doing what they would do is sometimes incompatible with conforming one's beliefs to the evidence. One complies with DN and ODN, but violates N itself (case (iv) in section 1). In Cohen and Comesaña's example, if you are the sort of person who usually conforms their beliefs to the evidence, then you may very well misjudge what conforming your beliefs to your evidence currently involves. For you are currently unaware of the perceptual illusion, and so may falsely take your current evidence to include matters about your environment of which you falsely take yourself to have perceptual knowledge. For the sake of argument, Cohen and Comesaña explicitly leave such externalist views of evidence open.²² On those views, one has less evidence in

the bad case than in the corresponding good case. Thus conforming one's beliefs to one's evidence in the bad case will generally involve something different from conforming one's beliefs to one's evidence in the good case, such as a difference in one's degree of confidence in a given proposition. Consequently, rationality in Cohen and Comesaña's sense will also require something different in the bad case from what it requires in the good case. Even granted that in the good case it is irrational for you not to believe that there is a red table before you (at least, if the question arises), it does not follow that in the bad case too it is irrational for you not to believe that there is a red table before you. One of the differences between the two cases in what rationality requires might be that in the good case, but not the bad case, rationality requires you to believe that there is a red table before you. We must take such possibilities seriously, not dismiss them without argument, with respect to a norm of rationality such as Cohen and Comesaña's, since it is *not* of type ODN. They fail to register this point, because they take no account of normative distinctions like those explained in section 1. More generally, once we see the difference between what rationality requires on a given occasion and what a rational person — someone with a general disposition to do what rationality requires — would do on that occasion, we should at least take seriously the idea that rationality never requires one to believe something false.²³

Of course, the preceding considerations do not by themselves constitute a positive case for knowledge as the standard for justified belief, rather than some weaker standard, such as high probability on the evidence, or even than some standard stronger than knowledge. The latter is not the only norm for belief that can predict our temptation to accept the line of thought (*) against itself. More theoretical considerations are needed to

choose between those norms. But once we realize how easily norms for belief which one violates in sceptical scenarios generate secondary and tertiary norms with which one complies in those scenarios, we should abandon the naïve idea that the normative status which the subject's beliefs share in the good and bad cases must be justification, as opposed to blamelessness.

A further advantage of the present framework is that it offers the prospect of unifying 'objective' and 'subjective' norms for belief. As most epistemologists accept, a false belief is somehow defective. In particular, the brain in the vat's false beliefs are defective. Their defect is not that they violate a norm with respect to which the good and bad cases are equal, for the subject's beliefs in the good case are by hypothesis not defective. If the brain in the vat's false beliefs are justified, then a second norm is needed to explain why their falsity is a defect. One then wonders what the relation is supposed to be between those two norms. By contrast, if one starts with a truth-entailing standard *J* for justified belief, one can explain directly what is wrong with false beliefs, as violations of that norm, while still having the resources to explain a corresponding derivative norm *ODJ*, which the subject's beliefs comply with equally in the good and bad cases. Of course, truth is itself a truth-entailing standard for justified belief, but not on its own a satisfying standard for belief. If the brain in the vat believes that there is a vat nearby, because it seems to itself to see one, the belief is true, and the brain also complies with the disposition-based norms derivative from the truth norm, since it has a general disposition to form true beliefs (in normal circumstances), but there is still something unsatisfactory about the belief. Knowledge is the natural candidate for a truth-entailing standard of justified belief, because it unifies objective and subjective norms for belief.²⁴

Notes

1 Nagel, San Juan, and Mar (2013) found a positive correlation between the simple truth-value of a belief and its perceived degree of justification, which they conjecturally attribute to a form of egocentric bias. Incidentally, in their study a large majority of subjects, independently of ethnicity and gender, denied that the belief constituted knowledge, in contrast to results reported in earlier studies such as Weinberg, Nichols, and Stich (2001).

2 See Alston 1989.

3 This brain was envatted only the night before, so semantic externalism allows it to have beliefs about hands, not just about patterns of neural activation.

4 The knowledge norm of assertion may also be overridden by weightier moral or prudential considerations (Williamson 2000, p. 256), even though no such exceptions are built into the content of the norm itself.

5 Read 'someone' here as taking wider scope than 'in S'. There are many potential complications. First, the class of persons may vary from one possible situation to another. Second, under indeterminism there may be no one thing that a given person would do in a given situation. Third, doing what *someone* who complies with N would do may be an

unhelpfully weak condition, because someone who would comply with N in enough possible situations to count as disposed to comply with N may nevertheless have a blip in the situation at hand and do almost anything. Perhaps most DN-compliers would comply with N in the circumstances, but a few would not. Strengthening ODN to rule out such cases may jeopardize the entailment from DN to ODN. Nevertheless, for most present purposes we can ignore these complications.

6 In the jurisprudential debate on justifications and excuses going back to Austin (1956-7), the issue of responsibility looms large. In respect of belief, that raises the difficult question whether belief is voluntary in any sense, which we can avoid for present purposes. For some contributions to the jurisprudential debate see Baron 2005, 2006; Duff 2006, Gardner 2007.

7 I am not suggesting that all excuses are of the kind described. That one was distraught with grief is an excuse of a quite different kind. Excuses are inexhaustibly various; one should not expect a neat taxonomy.

8 'Belief' here means flat-out belief, not just high credence. One can have extremely high credence that one's ticket will not win the lottery (as measured by one's betting behaviour, for instance), without believing flat-out that it will not win; one does not throw the ticket away. No good norm forbids high credence without flat-out belief in each member of an inconsistent set of propositions.

9 Direct reference theories count belief that Hesperus is bright and Phosphorus is not bright as belief in a contradiction (the conjunction of a proposition with its negation); one might prefer to restrict the consistency requirement to cases where the contradiction is believed under a guise of a form such as 'A & ¬A'. Dialetheists, who think that there are true contradictions, should adjust the requirement accordingly, replacing the contradictory form by one that entails everything by their lights. The point is not that explicit contradictions have a uniquely pathological status but that they serve to draw a clear line in a convenient place.

10 It is assumed here that knowledge entails belief.

11 The present concern is with doxastic rather than propositional justification.

12 '[I]n that sense of "justified" in which S's being justified in believing P is a necessary condition of S's knowing that P, it is possible for a person to be justified in believing a proposition that is in fact false' (Gettier 1963).

13 For instance, Cohen 1984, p. 280, argues that the conditions that justify a belief 'about the truth of some empirical proposition' do not entail the truth of that belief, because the evil demon can arrange for a justifying condition to obtain while the belief is false.

14 J and ODJ will still not coincide with DJ, for the simple reason that how one acts in the given situation determines whether one complies with J and ODJ but does not determine whether one complies with DJ. DJ is too global a norm to be a tenable standard of justified belief.

15 For the function DOD is plausibly equivalent to D, at least under simplifying assumptions. In other words, one is disposed to do what someone disposed to comply with a norm would do if and only if one is disposed to comply with that norm. Then for such a norm N, $OD(ODN) = O(DOD)N = ODN$. Here is a toy model in which DOD is equivalent to D. There is a fixed set of agents and a fixed set of situations. For any agent A and situation S, 'what A does in S' is well-defined. Situations are divided once and for all into two classes, the normal and the abnormal. For any norm N, if an agent complies with N in every normal situation, then she counts as complying with DN in every situation; otherwise she counts as complying with DN in no situation. An agent counts as complying with ODN in a situation if and only if what she does in that situation is what some agent who complies with DN does in that situation. Now consider a norm N that is *occurrent* in the sense that whether an agent complies with N in a situation is determined by what the agent does in that situation (when we are interested in a dispositional norm DN, it is natural but not mandatory to assume that N is *occurrent*). Suppose that A complies with DN. Then in any situation, what A does is what some agent who complies with DN does, so A complies with ODN. A fortiori, A complies with ODN in every normal situation, so A complies with DODN. Conversely, suppose that A complies with DODN. Let S be a normal situation. Then what A does in S is what some agent A* who

complies with DN does in S. Since N is an occurrent norm, A complies with N in S if and only if A* complies with N in S. But A* complies with N in S, because A* complies with DN and S is normal. Consequently, A complies with N in S. Thus A complies with N in every normal situation. Therefore A complies with DN. Thus in the model, $DN = DODN$ for every occurrent norm N. How far the result can be generalized is a matter for investigation. One can certainly construct not unreasonable models in which the equation fails even for some occurrent norms.

16 In the first Meditation, Descartes supposes that even in the most elementary reasoning, he may be confident that he is reasoning correctly when he is in fact reasoning fallaciously: Cartesian scepticism comprises scepticism about reason as well as scepticism about the external world.

17 See Martin 1994, Lewis 1997, and Bird 1998.

18 Clearly the norm applies only to flat-out beliefs, not to mere (high) degrees of confidence. I propose such a norm for belief at p. 47 of Williamson 2000: ‘Knowledge sets the standard of appropriateness for belief. [...] Mere believing is a kind of botched knowing’ (see also Williamson 1995, p. 563). For a recent discussion of knowledge as the norm of belief see Littlejohn 2013. In Williamson 2000, justification comes in degrees, measured by probabilities on one’s evidence, which is identified with one’s knowledge. For knowledge as the standard of *fully* justified belief see Williamson 2013a, pp. 11-12. Sutton 2007 defends the claim that justification requires knowledge. Bird

2007 defends a different knowledge-first account of justification; from the present perspective it would do better as an account of something subtly different from justification.

19 Conversely, Lasonen Aarnio 2010 uses a norm similar to DK to explain away nicely some supposed cases of knowledge defeat: in effect, she points out that a rash subject may comply with the K norm while violating both DK and ODK, but a confusion between the norms gives us the false impression that the subject violates K too. Compare also the derivation of secondary evidential norms for assertion from the primary norm of knowledge (Williamson 2000, p. 257; for a recent application see Benton 2013, pp. 357-8, drawing on DeRose 2002, p. 180, and 2009, pp. 94-5).

20 Cohen and Comesaña 2013b, p. 407. In a footnote they add a qualification: ‘One might hold that you are irrational only if you are taking some attitude toward whether there is a table’.

21 Cohen and Comesaña 2013a, p. 19.

22 On the equation $E = K$ of one’s total evidence with one’s total knowledge, defended in Williamson 2000, one misjudges one’s evidence in such cases. Cohen and Comesaña say ‘We are happy to grant $E = K$ for the sake of argument’ (2013b, p. 410).

23 Cohen and Comesaña (2013b, p. 407) complain that in Williamson 2000, like them and in contrast to Williamson 2013b, I rely without argument on the view that one can rationally believe a falsehood. In the passages they quote from Williamson 2000, I am using the word ‘rationally’ for complying with something like an ODN norm of doing what someone with good general cognitive dispositions would do; as those cases are most naturally envisaged, one complies with ODN because one complies with DN. By contrast, in Williamson 2013b I use the term for the non-ODN norm of conforming one’s belief to one’s evidence, in order to engage more directly with the statement in their previous paper ‘Rationality requires one to conform one’s beliefs to one’s evidence’ (2013a, p. 19). The terminological shift was confusing, and I should have warned the reader, but it did not signal any significant theoretical shift. Unlike ‘knowledge’, ‘rationality’ has never been one of my key theoretical terms, which is why I was so willing to adapt its use to the occasion in Williamson 2013b. In Williamson 2000, I argued for a view on which, *whatever* rationality is, we are sometimes rational without being in a position to know that we are being rational, and sometimes irrational without being in a position to know that we are being irrational, so that being in a position to know that one is being rational and not being in a position to know that one is being irrational are tempting alternative standards not equivalent to the original one, with a consequent destabilizing effect on the norm of rationality (see Srinivasan 2013a and 2013b for a defence of the underlying ‘anti-luminosity’ argument and further exploration of its normative consequences respectively). In effect, my concern was with a different style of derivative norm: for a norm N, there is the derivative norm KN of knowing that one is complying with N. While I still endorse the arguments in Williamson 2000, I now

want to do justice as well to the differently structured destabilizing effect of alternative disposition-based standards such as DJ and ODJ.

23 Thanks to audiences at the universities of Barcelona, Edinburgh, Geneva, Iowa, Luxembourg, Michigan, Olomouc, Oxford, St Andrews, Tel Aviv, the University of California at Riverside, MIT, Trinity College Dublin, and King's College London, and at conferences in Fortaleza and Xiamen, for helpful discussion of earlier versions of some of these ideas, and to Matthew Benton, David Enoch, Julien Dutant, Antti Kauppinen, Arturs Logins, and Christian Piller for detailed written comments on earlier versions of the material.

References

- Alston, William 1989: *Epistemic Justification*. Ithaca: Cornell University Press.
- Austin, J.L. 1956-7: 'A plea for excuses', *Proceedings of the Aristotelian Society*, 57: pp. 1-30.
- Baron, Marcia 2005: 'Justifications and excuses', *Ohio State Journal of Criminal Law*, 2, pp. 387-406.
- Baron, Marcia 2006: 'Excuses, excuses', *Criminal Law and Philosophy*, 1, pp. 21-39.
- Benton, Matthew 2013: 'Dubious objections from iterated conjunctions', *Philosophical Studies*, 162, pp. 355-358.
- Bird, Alexander 1998: 'Dispositions and antidotes', *The Philosophical Quarterly*, 48, pp. 227-234.
- Bird, Alexander 2007: 'Justified judging', *Philosophy and Phenomenological Research*, 74, pp. 81-110.
- Cohen, Stewart 1984: 'Justification and truth', *Philosophical Studies*, 46, pp. 279-295.
- Cohen, Stewart, and Comesaña, Juan 2013a: 'Williamson on Gettier cases and epistemic logic', *Inquiry*, 56, pp. 15-29.
- Cohen, Stewart, and Comesaña, Juan 2013b: 'Williamson on Gettier cases in epistemic logic and the knowledge norm for rational belief: a reply to a reply to a reply', *Inquiry*, 56, pp. 400-415.
- DeRose, Keith 2002: 'Assertion, knowledge, and context', *The Philosophical Review*, 111, pp. 167-203.
- DeRose, Keith 2009: *The Case for Contextualism*. Oxford: Clarendon Press.

- Duff, Antony 2006: 'Excuses, moral and legal: a commentary on Marcia Baron's "Excuses, Excuses"', *Criminal Law and Philosophy*, 1, pp. 49-55.
- Gardner, John 2007: 'The gist of excuses', in his *Offences and Defences: Selected Essays in the Philosophy of Criminal Law*, Oxford: Oxford University Press, pp. 121-139.
- Gettier, Edmund 1963: 'Is justified true belief knowledge?', *Analysis*, 23, pp. 121-123.
- Lasonen Aarnio, Maria 2010: 'Unreasonable knowledge', *Philosophical Perspectives*, 24: pp. 1-21.
- Lewis, David 1997: 'Finkish dispositions', *The Philosophical Quarterly*, 47, pp. 143-158.
- Littlejohn, Clayton 2013: 'The Russellian retreat', *Proceedings of the Aristotelian Society*, forthcoming.
- Martin, C.B. 1994: 'Dispositions and conditionals', *The Philosophical Quarterly*, 44, pp. 1-8.
- Nagel, Jennifer, Valerie San Juan, and Raymond Mar 2013: 'Lay denial of knowledge for justified true beliefs', *Cognition*, 129, pp. 652-661.
- Srinivasan, Amia 2013a: 'Are we luminous?', *Philosophy and Phenomenological Research*, forthcoming.
- Srinivasan, Amia 2013b: 'What's in a norm?'. Typescript.
- Sutton, Jonathan 2007: *Without Justification*, Cambridge, Mass.: MIT Press.
- Timmons, Mark, John Greco, and Alfred Mele (eds.) 2007: *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*, Oxford: Oxford University Press.
- Weinberg, Jonathan, Shaun Nichols, and Stephen Stich 2001: 'Normativity and epistemic

- Intuitions', *Philosophical Topics*, 29, pp. 429-460.
- Williamson, Timothy 1995: 'Is knowing a state of mind?', *Mind*, 104, pp. 533-565.
- Williamson, Timothy 2000: *Knowledge and its Limits*, Oxford: Oxford University Press.
- Williamson, Timothy 2007: 'On being justified in one's head', in Timmons, Greco, and Mele 2007, pp. 106-122.
- Williamson, Timothy 2013a: 'Gettier cases in epistemic logic', *Inquiry*, 56, pp. 1-14.
- Williamson, Timothy 2013b: 'Response to Cohen, Comesaña, Goodman, Nagel and, and Weatherson on Gettier cases in epistemic logic', *Inquiry*, 56, pp. 77-96.