Epistemological Ambivalence

(to appear in Nicholas Hughes (ed.), *Epistemic Dilemmas*, OUP)

Timothy Williamson

1. *Introduction*

A competent mathematician has just proved a surprising new theorem. She shows her proof to several distinguished senior colleagues, who all tell her that it involves a subtle fallacy. She cannot quite follow their explanations of her mistake. In fact, the only mistake is in their objections, obscured by sophisticated bluster; her proof is perfectly valid. But is she in a position to know that? What is she justified in believing?

Our mathematician is likely to feel *torn*, pulled in opposite directions. On one side, she has a clear sense of how her proof works. On the other, her colleagues' well-established expertise gives her strong reason to believe that her apparently clear sense must be some sort of intellectual illusion, a phenomenon not unheard of in mathematics. How can she do justice to both sources of evidence?

The mathematician's dilemma (if we may call it that) has an echo in the epistemology seminar room: what are *we* justified in believing about what *she* is justified in believing? We too may feel torn. But there are differences between the two perspectives. Although *we* can dismiss the possibility that the proof is invalid, since its validity is a stipulated condition of the example, it does not automatically follow that *she* can dismiss that possibility too. We can also complain, as she cannot, that in other respects the case is under-specified: the answers to our questions may depend on how complicated her proof is, how many colleagues she consulted, and how well-qualified they are.

Nevertheless, the differences in perspective between the mathematician and the epistemologist should not be exaggerated. The reflective mathematician asks epistemological questions about her own position: 'Am I justified in believing that my proof is valid?' Conversely, in considering what the mathematician is justified in believing, careful epistemologists imaginatively adopt her position, to take into account the limits of what is available to her. Although they are better off than she is with respect to the first-order question 'Is the proof valid?', since the answer is built into their description of the case, they may be no better off than she is with respect to the second-order question 'What is someone in the mathematician's position justified in believing about whether the proof is valid?'

In brief, if the mathematician believes that her proof *is* valid, she seems to disrespect the evidence of her colleagues' testimony. But if she believes that her proof is *not* valid, she seems to disrespect the evidence of the proof itself.

One way out is *agnosticism*. The mathematician may suspend judgment as to whether her proof is valid. The epistemologist may endorse or condemn her suspension of judgment, or suspend judgment on that higher-level issue. An attraction of agnosticism is that if one avoids judging, one is not guilty of misjudging.

On some views, belief comes in degrees of confidence, perhaps ideally conforming to the standard axioms of numerical probability theory. Then the questions become more nuanced. *How* confident should the mathematician be that her proof is valid? How confident should the epistemologist be that the mathematician's confidence that her proof is valid should lie in a given interval? Once degrees of confidence are on the menu, agnosticism is a less easy option, since middling degrees of confidence may be justified or unjustified, just like more extreme degrees. Perhaps one can avoid having any degree of confidence at all in a proposition by refusing to assess it, though one may be called on to justify even that refusal, just as one may be called on to justify suspension of belief.

Of course, none of this amounts to a demonstration that, whatever the mathematician does, she violates some epistemic duty. On any full specification of the case, for all that has been said so far, there *may* be a right attitude for her to take, involving no such violation. Yet, whatever the details, it is hard to escape a nagging feeling that if she accepts her proof as valid, she is not doing justice to her colleagues' testimony, while if she does not accept it as valid, she is not doing justice to the proof itself. The case has something like the phenomenology of a dilemma. It has that phenomenology for the mathematician herself: although she can see no flaw in the proof, she feels the genuine epistemic pressure to defer to her colleagues' expertise. It can also have that phenomenology for us, as epistemologists, participating in her struggle. Evidently, if the case is a dilemma, it is an epistemic one.[1]

Like paradigm moral dilemmas, the case concerns a responsible, self-conscious, language-using agent, fully aware of what is at stake and deliberately managing her own cognitive projects. Much epistemology considers only such agents. Yet epistemic phenomena are not nearly so confined. Small children and non-human animals have extensive knowledge of their environment, and an even wider range of true or false belief about it. Thinking about them is often a simpler and better place to start in epistemology than thinking about well-educated adult humans, with their more complex, self-reflective cognitive lives. Such cognitive sophistication is built on top of, and out of, animal cognition. Thus one may wonder whether apparent epistemic dilemmas are an artefact of the self-reflective, self-critical, deliberative part of sophisticated agents' cognitive activity, or instead are deeply rooted in more general features of cognition. If the latter, we still cannot expect all the phenomenology of a paradigm dilemma to generalize to small children and non-human animals. Rather, we should understand the stereotypical phenomenology as something like the explicit awareness in sufficiently sophisticated agents of underlying cognitive tensions which also occur implicitly in much more primitive agents too. The next section follows that approach.

2. *Primitive dilemmas*

In a paradigm dilemma, all options are somehow *bad*. Each involves violating a duty, or an obligation, or some other sort of exigent norm. But how can it be appropriate to apply such normative standards to young children or non-human animals? Is this not already a reason for restricting the category of dilemmas to responsible agents, who can be appropriately treated as answerable to normative standards?

Those rhetorical questions assume a very narrow reading of 'norm'. On a less restricted reading, all sorts of biological phenomenon are subject to broadly functional norms, with respect to which they can be classified as defective or non-defective. Something is *wrong* with a heart incapable of pumping blood. Something is *wrong* with a spider's web incapable of catching flies or other prey. Similarly, suppose that beliefs have a biological function which entails supplying the animal with information on which to act, just as hearts have the biological function of pumping blood, and spiders' webs have the biological function of catching flies and other prey.[2] Then something is *wrong* with a belief incapable of providing information on which to act; it is defective.[3]

Of course, there is no point—biological or not—in the animal being supplied with propositions selected at random, irrespective of truth-value, on which to act. It would not survive for long on that basis. Rather, information should be understood as consisting of *facts* or *true* propositions (which need not be encoded linguistically). Plausibly, beliefs have a biological function which entails supplying the animal with truths on which to act, in order to attain its goals.

None of that is to deny that false beliefs sometimes help animals towards their goals, if only by dumb luck. Self-overestimation may enhance the motivation to succeed—though it causes many failures too. Believing the same things as others facilitates being accepted as a member of the group, even if those beliefs are false—though conformity sometimes leads a group to disaster. Nevertheless, useful false beliefs are somehow deviant. Something's biological function is not just *whatever* it does that happens to benefit the animal. If someone gets rich—even predictably rich—by exhibiting his curiously deformed nose, that does not make it part of his nose's biological function. It has nothing to do with why he has a nose in the first place.

One obvious problem with conceiving biological function so generally as benefit to the animal (or indeed species) is that it assigns the *same* biological function to everything. Such an undifferentiated view of function would be of little use when we try to understand biological structure in terms of biological function. Productive functional theorizing employs much more specific functions than that.

Functions must not be *too* specific either. Some beliefs boost self-confidence, but others do the opposite, and most are neutral in that respect. Boosting self-confidence is not a *general* function of belief.

There is a standard way for belief to help the animal attain its goals. For example, the animal is thirsty. It believes that there is a pool of water by those trees, so that if it goes over there it will be able to drink. Consequently, it goes over to those trees. Since its belief is *true*, there *is* a pool of water by them. It drinks. Without the truth of the belief, such

explanations stall. Thus there is a natural point to restricting the function of belief to supplying *truths* or *facts* for the animal to act on. Since false beliefs cannot serve that function, they are defective.

Perhaps the function of belief is to supply something more than truth: in particular, *knowledge*. After all, knowledge implies *access* to the fact in a sense in which mere true belief does not. Whether the function of belief is to be knowledge depends on tricky theoretical issues about what it is to be a function. For present purposes, we only require the function of belief to involve *at least* truth. As will be seen below, a strong connection with knowledge can anyway be made from that starting-point.

There is a worry that some true beliefs lack practical consequences, and so will count as defective, because the animal cannot act on them. But could a belief really be impossible *in principle* to act on? For creatures who communicate by language, testimony can connect almost any belief to action. You believe *p*. Someone you trust tells you that if *p* holds and you open the door, you will get something you greatly value. You open the door. Admittedly, such testimony-based connections are not available to non-linguistic creatures, but their beliefs are presumably more restricted in topic, in ways more closely related to their practical interaction with their environment, so the problem of impractical beliefs is anyway less pressing for them.

Of course, many true beliefs are *unlikely* ever to be acted on, but that does not make them defective. Imagine a type of antibody in your blood whose function is to protect you against a specific type of lethal pathogen. That type of pathogen is so rare that you are unlikely ever to encounter one. But that does not make those antibodies *defective*. They are there for you if you need them. Similarly, true beliefs you are unlikely ever to act on are still there for you if you need them. Although *false* beliefs are also there for you if you need them, they are defective for a quite different reason: they are like antibodies which do not hinder the pathogen, or hinder it only by some deviant causal chain.

Thus, without treating young children and non-human animals as responsible agents, we can still classify their false beliefs (when they have them) as *defective*: something is wrong with them. They cannot serve the biological function of beliefs.

Confining the term ‘believe’ to biological believers may be too narrow. Perhaps intelligent robots can also have beliefs, since they are disposed to act on some propositions and not on others. We need not assume otherwise. If so, the robots’ beliefs may be true or false, and their false beliefs are defective too. They cannot serve the function of beliefs, if ‘function’ is confined to strictly biological function. This is not the place to address the large literature on what functions are, except to observe that, for understanding belief, it is more promising to generalize outwards from biological functions than from the *intended* functions of artefacts. Human beliefs are not like artefacts, deliberately designed.

Still, by itself, a norm for distinguishing between defective and non-defective beliefs cannot generate dilemmas. The reason is simple: it is vacuously satisfied in the absence of beliefs; there is nothing to be defective, by the standard of that norm. There agnosticism really does provide a way out. Admittedly, to be entirely without beliefs is hardly a feasible option, short of death, for animals which normally have beliefs: in waking hours, their perceptual systems produce streams of beliefs, almost automatically. Thus, on all available options, the animal has many beliefs. One might argue that this constraint combined with a

norm of non-defective belief could generate epistemic dilemmas: on each available option, the animal has some false beliefs. But, to show that, one would still have to argue that local agnosticism is not an available option, on which the creature has no beliefs at all about the matters at issue in the purported epistemic dilemma. Showing that might be hard. The threat of a dilemma is more powerful if driven by a further norm capable of positively requiring belief.

At first sight, such a positive belief norm seems to demand a different model of normativity. For *what* is supposed to be defective when the creature simply lacks relevant beliefs? By hypothesis, there is no belief to be defective. But that problem depends on too limited a view of cognition. After all, if the function of beliefs is to supply the animal with truths to act on, that function is not served when the animal has no (relevant) beliefs. Trivially, that is not a defect in any particular belief, but it *is* a defect in the animal's belief system as a whole, its system for generating beliefs, and in its relevant parts. False belief misguides action; lack of belief leaves it unguided. For purposes of action guidance, a balance must be struck between being too willing to believe, and thereby getting too many false beliefs, and not being willing enough to believe, and thereby not getting enough true beliefs. Those forces pushing in opposite directions may suffice to generate epistemic dilemmas.

However, putting the issue in terms of willingness or unwillingness to believe is not very helpful, since it leaves the connection with truth and falsity unexplained. For there is little point in striking a delicate balance between willingness and unwillingness to believe if the truth-values of the beliefs are distributed at random. Rather, the cognitive system must have *ways of producing* predominantly true beliefs. When everything goes right, the beliefs so produced are true. When something goes wrong, they may be true or false—or no beliefs may be produced.

The natural hypothesis is that when everything (or enough) goes right in the belief system, one has true beliefs *because one has knowledge*. When something (or too much) goes wrong, one lacks knowledge: at best one has true belief, at worst false belief, or none. Thus, when things go well, the animal acts on what it *knows*.[4] When things go badly, it acts on what it merely believes.

A natural development of this hypothesis is that the defective case, when things go wrong, is best understood in relation to the non-defective case, when they go right: thus mere belief is best understood in relation to knowledge. In a medical diagnosis of a disease, to understand what has gone wrong in the body, where and how, one needs to understand how it works when it is functioning normally. Similarly, to diagnose a defect in a belief, one needs to understand the contrast case of knowledge.

Of course, such programmatic big-picture remarks about the centrality of knowledge are highly contentious; they need detailed support. Some of it has been supplied elsewhere (Williamson 2000, 2017a), but much remains to be done. For the time being, I will just offer in support a schematic analogy, in the next section, this time with a phenomenon that is not specifically biological.

3. Cognitive systems and communication systems

Consider a simple *communication system*.[5] A source inputs a message somewhere in the system; when all goes well, the same message is output to the receiver somewhere else in the system. 'Source' and 'receiver' here are purely functional terms. Sources and receivers need not be agents; they may be biological cells or microchips. The message is *not* assumed to have propositional content; that would make communication systems too close to the cognitive systems we are trying to understand to serve as a helpful comparison. Thus messages need not be evaluable as true or false. The function of the communication system is simply to ensure the (qualitative) identity of the signal received with the signal sent, though what is identical may be something quite abstract, such as a pattern of sound waves. For present purposes, there is no further question as to whether it is 'true'. More complex communication systems may transform the signal in various ways, making the required relation between signal sent and signal received a one-one relation other than identity, but for illustrative purposes we can focus on the simpler case.

The identity of input and output is not automatic. There is *noise* in the system, which can interfere with the signal. As a result, the signal received sometimes differs from the signal sent. When the specific features of the signal received result entirely from the interference, and are causally independent of the signal sent, it is all *noise*. In the opposite case, when there is no interference, the signal received is all *message*. When there is partial interference, the signal received is part message, part noise, in a ratio dependent on the degree of interference. The distinction between message and noise in a signal received is *extrinsic*, since it depends on the signal's causal history, not just on its intrinsic structure. Thus the very same wave pattern could be all message, all noise, or part message, part noise, in various proportions, depending on its causal history.

The ratio of message to noise may vary, depending on external conditions. For example, there may be proportionately more noise during an electrical storm. When the ratio of message to noise is high, we describe the system as *working well*. When the ratio is low, we describe it as *working badly*.

Even when the system is working badly, the signal received may still happen to be (qualitatively) identical with the signal sent. Imagine lightning striking twice. After the first strike, the signal is changed utterly; it is all noise. After the second strike, the signal is changed utterly again; it is still noise—even if as a result the signal received chances to be (qualitatively) identical with the signal originally sent. Since it is all noise and no message, the system is still working badly.

The effect of a lightning strike may be either permanent or temporary. If it is permanent, the strike *damaged* the system. If the effect is only temporary, the strike left the system undamaged. Nevertheless, even in the latter case, the system was not working well during the interference. It may have been behaving just as systems of that type normally do behave in those circumstances, but then systems of that type are not made to work well while struck by lightning—that would be too much to expect.

Here is the proposed analogy between communication systems and cognitive systems. A signal received stands to the signal sent as a belief stands to the truth—more

specifically, in the latter case, as a state of believing the proposition *p* stands to the true one of *p* and its negation ¬*p*. Thus true belief corresponds to the (qualitative) identity of the signal received with the signal sent. This is only an analogy, because such identity does not make the signal received literally true, and the absence of such identity does not make the signal received literally false. The (qualitative) identity of the signal received with the signal sent is necessary but not sufficient for the signal received to be all message. Analogously, the truth of a belief is necessary but not sufficient for the belief to constitute knowledge. The distinction between message and noise is analogous to the distinction between knowledge and ignorance, *not* to the distinction between true and false belief. Just as the communication system is working well only when the message-to-noise ratio is high, not whenever the signal received is (qualitatively) identical with the signal sent (or at least very similar), so the cognitive system is working well only when the belief constitutes knowledge, not whenever the belief is true.

Like a lightning strike, a sceptical scenario need not involve structural damage to the system. Nevertheless, just as the absence of such damage does not mean that the communication system works well when lightning strikes, so it does not mean that the cognitive system works well in a sceptical scenario—that would be too much to expect.[6]

The analogy between communication and cognitive systems casts light on agnosticism as an escape from the threat of error. What corresponds to the absence of belief is a communication system in which no signals are received. *A fortiori*, no signal received differs from the signal sent. Needless to say, a communication system arranged not to receive signals is not working well—it is not working at all. For it to be working well, it must be open to receiving signals, because it must be open to receiving messages. Analogously, for the cognitive system to be working well, it must be open to acquiring beliefs, because it must be open to acquiring knowledge. Just as a communication system open to receiving messages has to run the risk of receiving instead signals different from those sent, so a cognitive system open to acquiring knowledge has to run the risk of acquiring instead false beliefs.

We can classify failures to work well as *local* or *global*. When a received signal just happens to be too noisy, it is a local failure, a defect in that particular signal. But when a communication system is too prone to receiving such noisy signals, it is a global failure, a defect in the system as a whole. When the system is too prone to not receiving signals at all, it is another global failure, another defect in the communication system as a whole. Similarly, when a belief fails to constitute knowledge, it is a local failure, a defect in that particular belief. But when a cognitive system is too prone to produce such beliefs short of knowledge, it is a global failure, a defect in the system as a whole. When the system is too prone to not producing beliefs at all, it is another global failure, another defect in the cognitive system as whole. For both kinds of system, any given case involves failures either at both local and global levels, or at neither, or at one without the other. As theorists, we may feel ambivalent about the mixed cases, of local failure without global failure, or of global failure without local failure.[7]

A more nuanced account places failures and defects on a spectrum from the most local to the most global, depending on how much of the system is involved. But contrasts

can still arise between levels at which the system is working well and levels at which it is working badly. Ambivalence may still be felt.

The next section explores such mixed cases for cognitive systems.

4. Ambivalence

Sceptical scenarios can be ordered on a spectrum from the most local to the ever more far-reaching. At the local end, one receives hoax testimony about some matter of particular fact, or is tricked by an isolated visual illusion. In more far-reaching scenarios, one is a brain in a vat.[8] But all the usual sceptical scenarios involve a contrast: in specific ways one is deceived, while in more general ways one still thinks like a rational creature. Insofar as one is deceived, one's beliefs fail to constitute knowledge. But insofar as one still thinks like a rational creature, one still exercises knowledge-conducive cognitive dispositions, though unfortunately in unfavourable conditions which prevent them from actually yielding knowledge. In that way, even the brain in a vat's cognitive catastrophe is far from complete; it is less than fully global. The point of a standard sceptical scenario is effectively to pit the two levels against each other. In the circumstances, cognition's working well at one level is incompatible with its working well at the other level.

Of course, we can also conceive far-reaching scenarios in which one is wildly irrational, while under the impression that one is perfectly rational. Indeed, we hardly need to imagine such cases, since news media provide ample evidence of actual ones. But for present purposes they are of less epistemological interest, precisely because they lack the dramatic tension between the two levels.

Vulnerability to many standard sceptical scenarios requires no special cognitive sophistication. Whatever species it comes from, a brain can in principle be put into a vat. Many perceptual illusions arise from low-level, more or less hardwired features of perceptual systems. No minded creature is too naïve to make mistakes, and to be deceived by appearances. Obviously, *some* sceptical scenarios require a cognitively sophisticated victim—for example, to be tricked by hoax testimony. But that feature is quite inessential to the general phenomenon.

Naturally, unreflective victims of sceptical scenarios feel no ambivalence. From their perspective, nothing seems abnormal or defective. The cognitive dissonance is discernible only by a thinker who reflects on the sceptical scenario from a third-personal perspective, for instance in imagination—even if their own situation happens to be exactly that of the scenario, they do not know it. From the outside, one can see the victim as doing badly, by being suckered and forming false beliefs, but also as doing well—or at least as well as one would do oneself—by doing their best in the circumstances.

Some epistemologists may feel no ambivalence about the sceptical scenario. The victim has justified false beliefs; so what? The familiarity of that category is comforting, but should not make us forget the original unsettling power of sceptical scenarios. They are called 'sceptical' because they prompt sceptical thoughts, even without the aid of philosophical training. The full-on sceptic exploits them to argue that we are not even *justified* in holding beliefs of the kind at issue, whether we are actually in the good case or

the bad one: we lack justification as well as knowledge. The sceptical argument may be unsound, but its seductive quality suggests an underlying unease with the category of justified false belief.

'Justified' is a normative term. Indeed, many epistemologists treat the distinction between justified and unjustified belief as *the* central normative distinction in epistemology. A justified belief is supposed to be *OK*. But how can a false belief be OK?

According to most sceptics and many non-sceptics, one has exactly the same evidence, as well as exactly the same beliefs, in the corresponding good and bad cases; the difference is purely external. Thus one's beliefs are just as well supported by one's evidence in the bad case as in the good case. In that sense, they are equally rational and equally justified in the two cases. Elsewhere, I have argued that such a view depends on an inadequate account of evidence (Williamson 2000). Without rehearsing those arguments here, I will briefly sketch some of the conclusions.

One of the many ways in which it is bad to be in the bad case is that one has less evidence than one seems to oneself to have—less evidence than one has in the good case. Consequently, one's beliefs are less well supported by one's evidence in the bad case than they are in the good case. In that sense, they are less rational and less justified in the bad case. A natural extension of this thought is that a belief is *fully* justified only if it constitutes knowledge. But that does not mean that one is a less rational *person* in the bad case than in the good case. One may have the same general dispositions in the two cases to have rational beliefs, beliefs well-supported by one's evidence, although the bad case is less favourable than the good case to exercising those dispositions successfully. In unfavourable circumstances, cognition's working well at the comparatively global level of general dispositions to believe makes it work badly at the comparatively local level of individual beliefs. That underlying tension may be reflected in ambivalent and unstable epistemological thinking.[9]

Similar tensions may arise even in epistemically favourable circumstances. A classic example is the Preface Paradox. A long scholarly monograph contains thousands of statements, each based on meticulous research and rigorous argument, checked and rechecked. In the preface, the author apologetically warns the reader that despite all the care taken, the book will inevitably contain errors. That statement in the preface is itself supported by overwhelming evidence from the history of scholarly monograph publishing, and anyway makes it inevitable that the book contains errors: for unless it is itself an error, the book contains errors. Either way, the book contains at least one error. Since the honest author believes every proposition stated in the book, at least one of those beliefs is false, and so not knowledge. Yet each belief is the product of strongly knowledge-conducive dispositions, in epistemically favourable circumstances.[10] To withhold belief from any of the propositions at issue would be to resist a knowledge-conducive disposition. In such cases, the interrelation of the propositions in play guarantees that cognition's working well at the global level of general dispositions to believe will make it work badly somewhere at the local level of individual beliefs.

There is also a converse phenomenon. In some circumstances, cognition's working badly at the global level of general dispositions to believe makes it work well at the local level of individual beliefs. That is what Maria Lasonen-Aarnio (2010) calls 'unreasonable

knowledge'. In such cases, global and local malfunctioning are again incompatible, but the trade-off between them goes the other way.

Our mathematician from section 1 may exemplify that phenomenon, on at least some fillings-in of the example. Suppose that she puts her colleagues' criticisms out of her mind, goes ahead and believes that her proof is valid, simply on the basis of her clear understanding of how it works. Perhaps, in that way, she can *know* that her proof is valid, despite the presence of what some epistemologists would regard as rebutting defeaters. Thus, at the level of that particular belief, cognition is working well. Nevertheless, at the level of general dispositions to believe, it is arguably working badly, given her tendency to ignore criticism from well-qualified sources. In many other cases, it will lead her astray—but not in this case. We have already explored the ambivalence and instability of epistemological thinking about that case.

A contrasting example of the phenomenon may be the thinker frozen by sceptical doubt in what happens to be a genuine sceptical scenario. The brain in a vat asks itself 'What if I am actually a brain in a vat?', and refuses to form the belief that it is not a brain in a vat. At the level of individual beliefs, it avoids forming a false belief; it thereby complies with the knowledge norm for belief. To that extent, locally cognition is working well. Nevertheless, globally, at the level of general dispositions to believe, it is working badly, given the brain's tendency to be frozen by extreme sceptical doubts. In its past history, the tendency may often have led it into a dead end, where it failed to acquire urgently needed knowledge. We may experience a similar ambivalence and instability in thinking about this case.

These are not simply examples of cognition working locally well but globally badly. They are cases where its working badly globally is the price to be paid for its working well locally. In the circumstances, working well locally is incompatible with working well globally. Our mathematician knows that her proof is valid only because she has a bad general tendency to ignore criticism from well-qualified sources. The sceptical brain in a vat avoids falsely believing that it is a brain in a vat only because it has a bad general tendency to be frozen by sceptical doubts. That is why such cases may be considered epistemic dilemmas.

5. *Local-local epistemic dilemmas*

Not all epistemic dilemmas directly depend on tension between norms on what one does and norms on what one is disposed to do. Some arise from tension between different norms on what one does.

We start with a schematic norm on ϕing, where for generality we make no assumption as to whether ϕing is epistemic or non-epistemic. When we apply our generic considerations to epistemic norms, the main question will be whether such norms somehow constitute an exception to the general rule.

Here is a norm on ϕing, where 'should' takes wide scope, as the brackets indicate, and 'C' expresses a condition:

N You should (ϕ if and only if C).

For example: you should water the plants if and only if they need it. Since 'should' distributes over conjunction, N is equivalent to the conjunction of its two directions:

NC You should (ϕ if C).
N¬C You should (not-ϕ if not-C).

In the previous example, NC says that you should water the plants if they need it, N¬C that you should not water the plants if they don't need it.

Naturally, agents have to apply N in light of their current evidence about their situation. It may not be transparent to you whether the plants need watering (C), and whether you watered the plants (ϕ). For present purposes, we can focus on uncertainty whether C, and assume for simplicity that there is no uncertainty whether you ϕ. To do well at complying with N, you typically need to be sensitive to evidence as to whether C. Thus implementing N naturally involves a derivative norm concerning some appropriate standard of evidence that C, for example, evidence that the plants need watering. Let us be more specific.

Given NC, in some sense it is wrong not to ϕ when C (not to water the plants when they need it). Then in a related sense it is also wrong not to ϕ when you have good evidence that C (not to water the plants when you have good evidence that they need it): you are failing to ϕ when you have good evidence that doing so involves violating the relevant norm NC, and so N. Thus N motivates a secondary norm ENC:

ENC You should (ϕ if you have good evidence that C).

The standard for 'good evidence' here is not intended to be very high; 80% probability on your evidence that C may suffice. Thus you can have good evidence for a false proposition. Even when the plants don't need watering, you can have good evidence that they do need it.

Similarly, given NC, in some sense it is wrong to ϕ when not-C (to water the plants when they don't need it). Then in a related sense it is also wrong to ϕ when you have good evidence that not-C (to water the plants when you have good evidence that they don't need it); you are going ahead and ϕing with good evidence that doing so involves violating the relevant norm N¬C, and so N. Thus N motivates another secondary norm EN¬C:

EN¬C You should (not-ϕ if you have good evidence that not-C).

'Good evidence' is to be understood in the same undemanding way in EN¬C as in ENC.

What is the status of 'derivative' norms like ENC and EN¬C? A natural worry is that ENC in effect competes with NC, and EN¬C with N¬C, for the same job. For both NC and ENC purport to give a deontically sufficient condition for ϕing, while both N¬C and EN¬C purport to give a deontically sufficient condition for not ϕing. Thus, it might be thought, proponents of NC and N¬C should be opponents of ENC and EN¬C, and *vice versa*.

In the example, what motivates the norm N and its two halves NC and N¬C is the plants' flourishing. Given that you complied with NC or N¬C, it makes no difference to the plants whether you also complied with or violated ENC or EN¬C. Equally, given that you violated NC or N¬C, it makes no difference to the plants whether you also violated or complied with ENC or EN¬C. In that sense, NC *screens off* ENC, and N¬C screens of EN¬C. Are derivative epistemic norms like ENC and EN¬C just an illusion?

The derivative epistemic norms come into their own when we assess the *agents* whom the original norm governs. Something went wrong with an agent who complied with NC and N¬C but violated ENC or EN¬C, even though the plants are flourishing. Equally, something went right with an agent who violated NC or N¬C but complied with ENC and EN¬C, even though the plants are wilting. Of course, what went right or wrong is only one detail in the agent's complex relation over time to the norm N, but it is a detail which counts in its own right. This is not restricted to reflective agents: N may be a functional norm for any creature with a cognitive life. Any such creature can be assessed with respect to its developing complex relation to N.

Violations of ENC or EN¬C are not simply indications of the agent's bad *intentions* concerning the original norm N. The agent may sincerely intend to water the plants if and only if they need it, but still lazily not water them while having good evidence that they need watering, or officiously water them while having good evidence that they don't need watering. Such failures matter in their own right. So do the agent's intentions, but the road to hell is paved with good intentions.

Nor are violations of ENC or EN¬C simply indications of the agent's bad *dispositions* concerning N. The agent may be generally disposed to water the plants if and only if they need it, but still on a particular occasion not water them while having good evidence that they need watering, or water them while having good evidence that they don't need watering. Such failures matter in their own right too. So do the agent's dispositions, but we care about performance as well as competence.

The converse mistake would be to go to the opposite extreme and treat ENC and EN¬C as displacing NC and N¬C. In the example, that would be to lose sight of the plants' flourishing as what motivates this normative structure in the first place. NC and N¬C *explain* ENC and EN¬C. Moreover, if NC and N¬C were displaced by ENC and EN¬C, then ENC and EN¬C would in turn be displaced by EENC and EEN¬C, which stand to ENC and EN¬C as the latter stand to NC and N¬C:

EENC You should (ϕ if you have good evidence that you have good evidence that C).

EEN¬C You should (not-ϕ if you have good evidence that you have good evidence that not-C).

An infinite sequence of such displacements reaches no equilibrium. One might hope to stop the infinite regress by declaring a fixed point, on the grounds that 'you have good evidence that you have good evidence that C' is just equivalent to 'you have good evidence that C' (likewise for 'not-C'). However, there are systematic objections to such principles about evidence (Williamson 2019). We may have to live with infinite hierarchies of derivative

epistemic norms. To keep them in perspective, we should remember that they cannot in general displace the original norm. The plant-watering example is a reminder of that.

The next task is to see how such derivative epistemic norms generate normative conflicts.

In themselves, the two evidential norms ENC and EN¬C are mutually compatible in their recommendations, given that you cannot simultaneously have both good evidence that C and good evidence that not-C. For we may assume that one has good evidence for a proposition only if it is more probable than not on one's evidence; thus, if one has good evidence for a proposition, one does not also have good evidence for its negation.

However, suppose that sometimes C even when you have good evidence that not-C. In such a situation, if you ϕ, you comply with NC, N¬C (and so N), and ENC, but violate EN¬C; if you do not ϕ, you comply with N¬C, ENC and EN¬C but violate NC (and so N). Either way, you violate one of the four norms.

Similarly, suppose that sometimes not-C even when you have good evidence that C. In such a situation, if you ϕ, you comply with NC, ENC, and EN¬C, but violate N¬C (and so N); if you do not ϕ, you comply with NC, N¬C (and so N), and ENC but violate EN¬C. Either way, as before, you violate one of the four norms.

Such conflicts between primary norms and their epistemic derivatives are not prototypical dilemmas, because the agent is by hypothesis not fully aware of the situation. It is the theorist, considering the conflict from a third-personal perspective, who is more likely to feel pulled both ways, in analysing the conflict. Nevertheless, we may call them 'dilemmas', because they instantiate the same basic normative structure.

Such epistemic dilemmas are impossible only if you cannot have good but misleading evidence as to whether C. In other words, both C/EC and ¬C/E¬C must hold in all relevant cases:

C/EC	If C, you lack good evidence that not-C.
¬C/E¬C	If not-C, you lack good evidence that C.

Thus you are vulnerable to epistemic dilemmas unless clashes are impossible between whether C and what your evidence indicates as to whether C.

Of course, if good evidence for a proposition entailed its truth, C/EC and ¬C/E¬C would immediately follow. But, as emphasized above, the intended standard for good evidence is much lower than that. The plants may need watering even though you have good evidence that they don't. In that case, complying with NC requires watering the plants, while complying with EN¬C requires not watering them. Equally, they may not need watering even though you have good evidence that they do. In that case, complying with N¬C requires not watering them, while complying with ENC requires watering them.

Do such conflicts arise when the original norm is itself epistemic? For example, we can interpret N as the norm that you should accept a proposition *p* if and only if you have good evidence for *p*. Then C/EC becomes the principle that if you have good evidence for *p*, you lack good evidence that you lack good evidence for *p*; ¬C/E¬C becomes the principle that if you lack good evidence for *p*, you lack good evidence that you have good evidence for *p*. Naturally, both principles hold if the status of a proposition on your evidence is always

epistemically transparent to you, and so built into your evidence. However, there is good reason to doubt that agents' evidence always is epistemically transparent to them (Williamson 2000, 2019, 2020). For example, it may be hard to know whether complex statistical data confirm or disconfirm a given hypothesis about a virus. The non-transparency of evidence was already noted in connection with the difference between ENC and EN¬C on one hand and EENC and EEN¬C on the other.

Admittedly, C/EC and ¬C/E¬C together do not entail the full transparency of evidence to the agent. However, once one rejects the picture of evidence as fully transparent, what alternative reason is there to endorse C/EC and ¬C/E¬C? For instance, suppose that your evidence appears to include a proposition *e* but does not really include *e*, and that, without *e*, you have good evidence for a proposition *p*, but with *e*, you lack good evidence for *p*. Since you lack *e*, you have good evidence for *p*; since you appear to have *e*, you also have good evidence that you lack good evidence for *p*. Thus C/EC fails on the envisaged interpretation. Similarly, suppose that, without *e*, you lack good evidence for a proposition *q*, but with *e*, you have good evidence for *q*. Since you lack *e*, you lack good evidence for *q*; since you appear to have *e*, you have good evidence that you have good evidence for *q*. Thus ¬C/E¬C fails on the envisaged interpretation.

Obviously, such scenarios should be developed in more precise detail to be fully convincing, but in general the prospects for C/EC and ¬C/E¬C are not bright on non-trivial epistemic interpretations. Technical results indicate that the gap between evidence and evidence about evidence can be very wide indeed. For instance, in some plausible epistemic models, one's evidence includes a given proposition even though it is almost certain on one's evidence that one's evidence does *not* include that proposition (Williamson 2014). These matters are of course controversial; I will not repeat detailed arguments already made elsewhere. What I intend here is just some exploration of the landscape and its consequences for epistemic dilemmas.

Consider a counterexample to C/EC on an interpretation of N as a general epistemic norm. You have good evidence for *p*, but you also have good evidence that you lack good evidence for *p*. In this case, complying with NC requires accepting *p*, while complying with EN¬C requires not accepting *p*. Thus the two epistemic norms conflict.

Consider a counterexample to ¬C/E¬C on the same interpretation of N. You lack good evidence for *q*, but you have good evidence that you have good evidence for *q*. In this case, complying with N¬C requires not accepting *q*, while complying with ENC requires accepting *q*. Again, the two epistemic norms conflict.

One feature of the norm N is that it leaves agents no discretion on whether to ϕ. That is determined by whether C. Thus one might hope to avoid epistemic conflicts by cutting agents some epistemic slack and leaving a buffer zone in which they are left to their own discretion on whether to ϕ. Thus one might replace N by a bipartite norm of NC and ND, where NC is as before but 'C' and 'D' are contraries rather than contradictories:

NC	You should (ϕ if C).
ND	You should (not-ϕ if D).

If ‘C’ and ‘D’ were contradictories, ND would be equivalent to N¬C, so the conjunction of NC and ND would just amount to N again. The intention now is to avoid that. For instance, if ‘C’ is ‘you have good evidence for *p*’, ‘D’ might be ‘you have bad evidence for *p*’ (*p* is not highly probable on your total evidence). In the intermediate zone between the two conditions, your evidence for *p* is neither good enough to mandate accepting *p* nor bad enough to forbid accepting it. Then NC and ND require nothing further of you; it is up to you whether you accept *p*.

As before, implementing NC and ND in light of your evidence naturally gives rise to corresponding derivative norms:

ENC You should (ϕ if you have good evidence that C).
END You should (not-ϕ if you have good evidence that D).

This combination gives agents some leeway even when C and D are not epistemically transparent:

If C (so not-D) but you have good evidence that not-C, you can still comply with all four norms NC, ND, ENC, and END by ϕing, provided that you lack good evidence that D.

Similarly, if not-C but you have good evidence that C (so you lack good evidence that D), you can still comply with all four norms by ϕing, provided that not-D.

If D (so not-C) but you have good evidence that not-D, you can still comply with all four norms by not ϕing, provided that you lack good evidence that C.

If not-D but you have good evidence that D (so you lack good evidence that C), you can still comply with all four norms by not ϕing, provided that not-C.

However, these norms still generate epistemic conflicts if C but you have good evidence that D. For in such a situation, if you ϕ, you violate END; if you do not ϕ, you violate NC. Equivalently, you comply with both NC and END only if C/¬ED holds:

C/¬ED If C, you lack good evidence that D.

For instance, if ‘C’ is ‘you have good evidence for *p*’, and ‘D’ ‘you have bad evidence for *p*’, C/¬ED says that if you have good evidence for *p*, you lack good evidence that you have bad evidence for *p*.

The norms can also generate epistemic conflicts if D but you have good evidence that C. For in such a situation, if you ϕ, you violate ENC; if you do not ϕ, you violate ND. Equivalently, you comply with both ND and ENC only if D/¬EC holds:

D/¬EC If D, you lack good evidence that C.

On the same values for ‘C’ and ‘D’ as before, D/¬EC says that if you have bad evidence for *p*, you lack good evidence that you have bad evidence for *p*.

Defenders of the doubly bipartite approach may hope to avoid dilemma-like combinations by picking ‘C’ and ‘D’ far apart, separated by a wide buffer zone. But the prospects for such a strategy are poor. It depends on the assumption that there is an adequate stock of conditions incapable of ever failing epistemic transparency too badly. As

already noted, there is little reason to accept that assumption, once the obstacles to full epistemic transparency have been recognized.

One might get the impression that complying with the evidential derivative of a norm is equivalent to doing as someone disposed to comply with the original norm would do, so that the local-local dilemmas discussed in this section are really just disguised versions of the local-global dilemmas discussed earlier. But it is not so. For if evidence is not epistemically transparent, no more is evidence about evidence. In epistemically unfavourable circumstances, good evidence that C may seem like good evidence that D, so that someone disposed to comply with NC and ND will thereby refrain from accepting *p*, intending to comply with ND, but thereby violating ENC. In the same circumstances, someone who accepts *p*, and thereby complies with ENC, will typically lack the disposition to comply with NC, and will not be doing as someone with the disposition to comply with NC would do. After all, if dispositions were unerring, someone with the disposition to comply with NC and ND would always comply with them, rather than with ENC and END. Of course, these schematic points need to be tested in detail, but they are very much in line with other considerations in this paper. Dispositions and evidence are just constituted too differently to march in lock-step. Local-local dilemmas are not local-global dilemmas in disguise.

6. Morals

Epistemic dilemmas are not simply curiosities to collect. Neglecting them can lead to large-scale distortions in both epistemology and normative theory. For consider two *correct* norms with the forms of NC and ND.

Suppose that C and you ϕ but have good evidence that D. Thus you comply with NC and ND, the norms stipulated to be correct. However, you violate END, a derivative norm motivated by the correct norm ND, so you may well be felt to be not doing as you should in ϕing. Consequently, NC is liable to *seem* incorrect, since it is the norm requiring you to ϕ. Such cases may be hailed as counterexamples to NC, even though NC is by hypothesis in fact correct.

Similarly, suppose that D and you do not ϕ but you have good evidence that C. Thus you again comply with NC and ND, the norms stipulated to be correct. However, you violate ENC, a derivative norm motivated by the correct norm NC, so you may well be felt not to be doing as you should in not ϕing. Consequently, ND is liable to *seem* incorrect, since it is the norm requiring you not to ϕ. Such cases may be hailed as counterexamples to ND, even though ND is by hypothesis in fact correct.

Thus dilemmas, including epistemic dilemmas, can lure us into rejecting correct norms.[11] Contrapositively, holding onto the correct norms can lure us into rejecting the possibility of such dilemmas, and so accepting problematic epistemological principles like C/¬ED and D/¬EC.

If something like the view sketched in sections 1-4 is accepted, there is much less pressure to adopt dubious epistemological principles in order to avoid epistemic dilemmas. For one has already conceded that epistemic dilemmas do occur, and so is already committed to making whatever adjustments in epistemology are needed to accommodate

them. The theoretical cost of acknowledging some further epistemic dilemmas of a somewhat different type is much lower. Of course, a parallel argument runs in reverse, from accepting the type of epistemic dilemma in section 5 to accepting the type in sections 1-4.

A more general moral is the need for greater caution in using our pre-theoretic judgments on particular cases in epistemology. We cannot do without reliance on them altogether, for without such specifically epistemological constraints, epistemological theory is liable—in practice as well as in principle—to float free of the reality which it is supposed to theorize about. But in cases of ambivalence, simply giving priority to one intellectual pull over another is unlikely to yield much insight. For a start, which way we jump may depend on the epistemologist's skill in describing the example so as to make us see it with the right *gestalt*. The psychological lessons to be learned from a duck-rabbit image do not depend on making the right decision as to whether to give priority to the duck, or to the rabbit. Rather, they involve understanding how both aspects can be potentially present in the same picture. We must be willing to give similarly nuanced analyses of examples in epistemology. But that is no licence to handle such cases unsystematically. Instead, we need a clear and effective theoretical framework within which to resolve the resultant forces into their components. The theory itself should be as simple and general as possible, with a minimum of moving parts, compatibly with having the resources to analyse the complexity of specific cases, and in particular to identify both horns of a dilemma.[12]

The reader may have noticed how little of the argument of the last two sections is specific to purely epistemic dilemmas. Section 5 explained how just about any norm generates derivative epistemic norms with which it sometimes conflicts. The hard part of the argument—only gestured at in this chapter—is showing that epistemic primary norms are no exceptions to the rule. Consequently, the methodological morals drawn in this section have a much wider potential application to normative theory.

Notes

1 See Hughes 20XX for the idea of an epistemic dilemma.

2 See Millikan 1984a and 1993 for an account of biological functions and biological normativity, and Millikan 1984b for its application to belief.

3 The word 'believe' and the more ordinary word 'think' are sometimes used for propositional attitudes close to *guessing* (Hawthorne, Rothschild, and Spectre 2016, Rothschild 2019). In guessing *p*, one need not be disposed to act on *p*. One can regard *p* as the most probable answer to a question, and even regard *p* as highly probable while unwilling to rely on *p* itself as an assumption—although one may be willing to rely on the assumption that *p* is highly probable (example: let *p* be the proposition that this ticket will not win the lottery). For philosophy, an attitude of belief more closely tied to action is of more interest. It is sometimes called 'outright belief'. This paper follows the dominant philosophical tradition by using the term 'belief' for outright belief. See Williamson 202Y for more discussion.

4 In Millikan's normative biological sense of 'Normal', beliefs are Normally knowledge; see Millikan 1984b. For a knowledge norm on the premises of practical reasoning see Hawthorne and Stanley 2008 and Williamson 2017a. For another biological account of knowledge see Kornblith 2002.

5 The *locus classicus* for the application of communication theory to epistemology is of course Dretske 1981.

6 The contentious idea that nearby fake barns make your true belief that there is a barn in front of you not knowledge (because you are lucky not to be facing a fake barn) corresponds to the contentious idea that nearby lightning strikes make the signal received not message (because the system was lucky not to be struck by lightning, even if the signal received is in fact (qualitatively) identical with the signal sent.

7 One difference between the two kinds of system is that communication systems (at least of the simple sort described) have no discretion in *which* signals to transmit, whereas cognitive systems have significant discretion in what questions are addressed and so in which propositions are current candidates for belief (even if the discretion is exercised at an unconscious level). Such discretion can be used well or badly, since some questions will be better than others on the relevant dimension, whatever that is. But this paper does not pursue the resultant normative issues.

8 To finesse Putnam's objection that the brain does not mean what we mean by 'brain in a vat', we may stipulate that it was only very recently envatted.

9 See Williamson 2017b and 202X for a detailed application of this idea to justification and rationality.

10 This is a significant difference between the Preface Paradox and the Lottery Paradox, given the plausible assumption that the belief that a given ticket will lose, based only on the number of tickets in the lottery, does not constitute knowledge.

11 See Srinivasan 2015 for more discussion of how epistemological errors distort normative theory.

12 Thanks to Nick Hughes and Daniel Kodsi for detailed constructive comments and participants in a class at Oxford for discussion on earlier versions of this material.

References

Dretske, Fred. 1981: *Knowledge and the Flow of Information*. Oxford: Blackwell.

Hawthorne, John, and Stanley, Jason. 2008: 'Knowledge and action'. *Journal of Philosophy*, 105: 571-590.

Hawthorne, John, Rothschild, Daniel, and Spectre, Levi. 2016: 'Belief is weak', *Philosophical Studies*, 173: 1393-1404.

Hughes, Nicholas. 20XX: 'Dilemmic epistemology', *Synthese*, forthcoming.

Kornblith, Hilary. 2002: *Knowledge and its Place in Nature*. Oxford: Oxford University Press.

Lasonen-Aarnio, Maria. 2010: 'Unreasonable knowledge', *Philosophical Perspectives*, 24: 1-21.

Millikan, Ruth Garrett. 1984a: *Language, Thought, and Other Biological Categories*. Cambridge, Mass.: MIT Press.

Millikan, Ruth Garrett. 1984b: 'Naturalist reflections on knowledge', *Pacific Philosophical Quarterly*, 65: 315-334. Reprinted in Millikan 1993.

Millikan, Ruth Garrett. 1993: *White Queen Psychology and Other Essays for Alice*, Cambridge, Mass.: MIT Press.

Rothschild, Daniel. 2019: 'What it takes to believe', *Philosophical Studies*, online.

Srinivasan, Amia. 2015: 'Normativity without Cartesian privilege', *Philosophical Issues*, 25: 273-299.

Williamson, Timothy. 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, Timothy. 2014: 'Very improbable knowing', *Erkenntnis*, 79: 971-999.

Williamson, Timothy. 2017a: 'Acting on knowledge', in Adam Carter, Emma Gordon, and Benjamin Jarvis (eds.): *Knowledge First: Approaches in Epistemology and Mind*, 163-181. Oxford: Oxford University Press.

Williamson, Timothy. 2017b: 'Ambiguous rationality', *Episteme*, 14: 263-274.

Williamson, Timothy. 2019: 'Evidence of evidence in epistemic logic', in Mattias Skipper and Asbjørn Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, 265-297. Oxford: Oxford University Press.

Williamson, Timothy. 2020: 'The KK principle and rotational symmetry', *Analytic Philosophy*, forthcoming.

Williamson, Timothy. 202X: 'Justifications, excuses, and sceptical scenarios', in Julien Dutant and Fabien Dorsch (eds.), *The New Evil Demon*, Oxford University Press, to appear.

Williamson, Timothy. 202Y: 'Knowledge, credence, and strength of belief', in Amy Plantree and Baron Reed (eds.), *Expansive Epistemology: Norms, Action, and the Social World*. London: Routledge, to appear.