PART I

The Debate between Conciliatory and Steadfast Theorists

A. Steadfastness

1

Disagreement Without Transparency
Some Bleak Thoughts[1]

John Hawthorne and Amia Srinivasan

1 Dashed hopes

The question at the centre of this volume is: *what ought one to do, epistemically speaking, when faced with a disagreement?* Faced with this question, one naturally hopes for an answer that is principled, general, and intuitively satisfying. We want to argue that this is a vain hope. Our claim is that a satisfying answer will prove elusive because of *non-transparency*: that there is no condition such that we are always in a position to know whether it obtains. When we take seriously that there is nothing, including our own minds, to which we have assured access, the familiar project of formulating epistemic norms is destabilized. In this paper, we will show how this plays out in the special case of disagreement. But we believe that a larger lesson can ultimately be extracted from our discussion: namely, that non-transparency threatens our hope for fully satisfying epistemic norms in general.

To explore how non-transparency limits our prospects for formulating a satisfying disagreement norm, we will put forward what we call the *Knowledge Disagreement Norm* (KDN). This norm falls out of a broadly knowledge-centric epistemology: that is, an epistemology that maintains that knowledge is the *telos* of our epistemic activity. We will then explore the ways in which KDN might be thought to be defective. In particular, we will explore the ways in which KDN fails to satisfy some common *normative* or *evaluative* intuitions. We will show, in turn, how this failure is a result of the non-transparency of knowledge: that is, the fact that one is not always in a position to know whether oneself, or someone else, knows a given proposition. We will then argue that this kind of failure is inescapable, as any plausible epistemic norm will feature a non-transparent condition. We

will conclude with some tentative remarks about what this might mean for the disagreement debate in particular and the project of formulating satisfying epistemic norms more generally. Ultimately we leave it to the reader to decide how bleak our case really is.

2 Some semantic preliminaries

When addressing the question 'What ought one to do, epistemically speaking, in the face of disagreement?' it can be useful to reflect on the semantics of ought-claims, and, in particular, the way in which ought-claims are notoriously context-sensitive. For 'ought' is one of a family of modals that, while retaining a barebones logical structure across its uses, is flexible in its contribution to the meaning of sentences. For example, there is a use of 'ought' connected to particular desires or ends, as in: 'The burglar ought to use the back door, since it's unlocked.'[2] There is also a use connected to legal or moral norms, as in: 'You ought to go to jail if you murder someone.'[3] And there is a use that is (arguably) connected to the evidential situation of an individual or group, as in: 'He ought to be in London by now.'[4]

Even within any one of these broad categories, there is considerable scope for context-dependence. For example, the truth conditions of a deontic ought-claim will also be sensitive to which facts are held fixed in the conversational context. For example, when one says of a criminal 'he ought to have gone to jail', one is holding fixed the fact of the crime. By contrast, when one says 'he ought to have never started on a life of crime', one isn't holding fixed the fact of the crime. Indeed it is plausible to suppose—and contemporary semantic wisdom does indeed suppose—that the semantic contribution of 'ought' is in general contextually sensitive to both a relevant domain of situations (the 'modal base'), and to a relevant mode of ranking those situations ('the ordering source').[5] On a popular and plausible version of this account, 'It ought to be that p' is true relative to a domain of situations and a mode of ordering just in case there is some situation in the domain such that p holds at it and at all situations ranked equal to or higher than it. Where there are only finitely many situations, this is equivalent to the constraint that p holds at all of the best situations. Note that this toy semantics has the consequence that finite closure holds for ought-claims. If 'It ought to be that p_1....ought to be that p_n' is true for a finite set of premises, then 'It ought to be that q' is true for any proposition

---

[2] This use is sometimes called 'bouletic'.

[3] This use is typically called 'deontic'.

[4] This is the so-called 'epistemic' use. For further discussion see John Hawthorne, 'The Epistemic Ought' (in progress). The above taxonomy is not intended to be canonical or exhaustive.

[5] The more technical terminology is due to Angelika Kratzer, who has done a great deal to promote the picture wherein modals like 'can', 'ought', and 'must' have an invariant core logical structure across their uses, and where a semantic value gets set by domain and mode of ordering. See notably Angelika Kratzer, 'What "Must" and "Can" Must and Can Mean', *Linguistics and Philosophy* 1 (1977): 337–55. See also Kai von Fintel and Sabine Iatridou, 'Anatomy of a Modal Construction', *Linguistic Inquiry* 38 (3) (2007): 445–83.

entailed by that set.[6, 7] This style of semantics is by no means sacrosanct, but it is one that will be in the background of our thinking in this paper. From this perspective, the question 'What ought one to do when one encounters a disagreement?' can thus be clarified by considering how the key contextual factors—ordering source and modal base—are being resolved at the context at which the question is being raised.

3 The Knowledge Disagreement Norm

From the perspective of a knowledge-centric epistemology—that is, an epistemology that takes the most central goal of our epistemic activity to be knowledge—it is natural to rank outcomes with knowledge over outcomes of withholding belief, which are in turn ranked over outcomes of knowledge-less belief:[8]

The Knowledge Disagreement Norm (KDN): In a case of disagreement about whether p, where S believes that p and H believes that not-p:

  (i) S ought to trust H and believe that not-p iff were S to trust H, this would result in S's knowing not-p[9]

---

[6] Note that within the framework of this semantics, apparent counterexamples to finite closure will be handled by appeal to context-shift. 'It ought to be that some murderers are never released from jail' sounds true, while 'It ought to be that there are some murderers' sounds false, and yet the that-clause of the former claim entails that of the latter. The standard resolution of this problem within the framework presented here says that contexts in which one says 'It ought to be that there are some murderers' would almost inevitably take a modal base that includes some non-murderer worlds (rendering the claim false), but contexts in which one utters the first sentence tend to hold fixed many more facts, including that there are some murderers.

[7] There will however be counterexamples to countable closure, the principle that extends the closure principle to any countable set of premises. Relevant here are infinitary normative dilemmas—that is, cases in which whatever one does of an infinite range of actions, one does something that one ought not to do. Suppose God allows one to fill out a blank cheque to a deserving charity with any natural number in pounds sterling. And suppose the relevant ranking is by size of gift; the bigger the gift, the better. One ought to write *some* amount down since situations in which one does this clearly outrank the situation in which one writes nothing down. But one ought not give exactly one pound since there are situations where one gives more than one pound that outrank any situation where one gives one pound. And one ought not give exactly two pounds for analogous reasons. . . . Hence anything one does is something that one ought not to do. Here we can generate a counterexample to countable closure by noting that the countable set of premises of the form 'It ought to be that one does not give exactly N pounds' (where N is a natural number greater or equal to 1) entails that it ought to be that one does not give some non-zero natural number of pounds.

[8] We are operating with an altogether natural gloss on 'S and H disagree about whether p', where this requires that S believe p and H believe not-p. At least as we ordinarily use the term, it is less natural to describe a case in which S believes p and H is agnostic about p as a case in which S 'disagrees' with H as to whether p.

[9] We might want to screen off cases in which the closest world in which one trusts H, the truth-value of p is different to the truth-value of p at the actual world. (In essence, this would involve restricting the modal base to worlds that match the actual world with regard to the truth-value of p). We can do this with a slightly more complicated counterfactual: S ought to trust H that not-p iff were S to trust H, and p have the truth-value it actually does, this would result in S's knowing not-p. Alternatively, we might want to complicate the test as follows: S ought to trust H iff (i) were S to trust H, S would come to know not-p, and (ii) were S to stick to her guns, S wouldn't come to know p. A test case: Fred says to you 'You believe everything I say'. You actually don't believe this claim, and for *that reason* the claim isn't true; however, if you were to believe the claim, it would be both true and known by you . . . For the sake of simplicity we ignore these interesting complications. Thanks to Cian Dorr here.

(ii) S ought to dismiss H and continue to believe that *p* iff were S to stick to her guns this would result in S's knowing *p*, and

(iii) in all other cases, S ought to suspend judgment about whether *p*.[10]

According to KDN, one should be 'conciliatory' in the face of disagreement—that is, give up one's belief that *p* and trust one's disagreeing interlocutor that not-*p*—just in case so trusting would lead one to know that not-*p*. Since (it is generally granted) trusting someone who knows is a method of acquiring knowledge oneself, (i) recommends that S trust H in cases where H *knows* that not-*p*.[11] Being conciliatory in such cases will lead S to greater knowledge.[12] According to KDN, one should be 'dogmatic' in the face of disagreement—that is, dismiss one's interlocutor and continue to believe *p*—if one knows that *p*. What about disagreement cases where neither S nor H knows whether *p*? In such a case, KDN demands that S suspend judgment.[13]

There are a few things worth noting from the outset about KDN. First, KDN is inspired by knowledge-centric epistemology, an epistemology that takes the *telos* of belief and epistemic practice more generally to be knowledge.[14] One might have, by contrast, a

---

[10] By suspending judgment we mean that one doesn't flat out believe the proposition or its negation. This is compatible with being very confident that a proposition is true or false; for example, one may be very confident that one is going to lose a lottery without flat out believing that one will.

[11] We ignore complications having to do with cases where even though H knows *p*, H wouldn't know *p* were S to trust H.

[12] Of course, there are cases in which trusting someone who knows might not result in knowing oneself—and certain kinds of disagreement might produce such cases. We discuss a view on which disagreement (or a subset of disagreements) 'defeats' the knowledge-transferring power of trust shortly.

[13] Tying ought-claims to counterfactuals gives contestable results in some cases. Suppose S is asked to donate some small amount of money to charity, and that it would be best (morally speaking) for S to donate the money and then continue on with her life as usual. But suppose further that, if S donated the money, she would fly into a rage and go on a killing spree. Insofar as we test for the truth of 'S ought to do A' by looking at the closest world where S does A, it is then true that S ought to make the donation and return to life as normal, but *not* true that S ought to make the donation. Thus, tying ought-claims to counterfactuals runs the risk of violating the principle that if S ought to A and B, then S ought to A. (For a possible fix see the next footnote).

[14] An attachment to a knowledge-centric epistemology need not lead us in a forced march to KDN. KDN ranks actions according to their counterfactual outcomes. But insofar as the domain of possibilities relevant to a normative claim concerning an action type involves more than the closest world where the action type is performed, there will be natural modes of ordering that violate KDN. Imagine, for example, that S currently knows *p*, and that if she were to continue to believe *p* *for her current reasons*, she would continue to know *p*; however, if S were to stick to her guns with regard to *p*, her basis for believing *p* would, as it happens, change to poor reasons, such that she would no longer know *p*. If the domain of worlds getting ranked includes worlds where the basis does not change, 'S ought to stick to her guns and not change her basis' will be true, and (recall finite closure) so will 'S ought to stick to her guns', even though the counterfactual 'If S were to stick to her guns she would maintain knowledge' is false. One can improve on KDN with the following amendment: one ought to stick to one's guns *iff* there is some way (where context provides the relevant space of ways of acting) of sticking to one's guns such that were one to stick to one's guns in that way one would maintain knowledge (and similarly for the other clauses). Take the person who believes *p* for bad reasons but has a good basis available. In a context in which the modal base includes worlds where the basis is switched, the claim 'S ought to stick to her guns' will come out true. In contexts where the subject's bad basis is held fixed, the ought-claim will come out false. This kind of context-dependence is exactly what one should expect. We do not dispute the need for such a refinement of KDN, and perhaps no counterfactual formulation is in the end problem-free. But as our discussion does not turn on the letter of KDN, we take the liberty of ignoring the needed refinements in the main text.

justification-centric epistemology, according to which the *telos* of belief is mere justified belief, or a truth-centric epistemology, according to which the *telos* of belief is mere true belief. Each of these alternative views could lead to disagreement norms that are analogous to the knowledge-centric KDN. While we will not discuss these possibilities here, much of the discussion that follows applies to them as well.

Second, KDN says nothing about how S and H should respond in cases where their disagreement is a matter of divergent credences as opposed to conflicts of all-or-nothing belief. Despite the prevailing trend in the disagreement debate, our discussion will for the most part proceed without mention of credences. We find this a natural starting point; our pre-theoretic grip on the phenomenon of disagreement tends to be in terms of conflicts of all-or-nothing belief.

Third, note that compliance with KDN will not necessarily result in overall knowledge maximization. Suppose S disagrees with H about whether Jack and Jill went up the hill together. If H were to trust S, H would come to know that Jack and Jill indeed went up the hill together, but he would also abductively come to believe a cluster of false propositions based on the (false) hypothesis that Jack and Jill are having an affair. In short, KDN is locally consequentialist with respect to the *telos* of knowledge. Less local consequentialisms are of course also possible, and we shall return to this issue in due course.

Fourth, KDN's gesture towards knowledge as epistemic *telos* can be unpacked in various ways, corresponding to different meta-epistemological views. On one gloss, the relevant 'ought' is bouletic/desire-based: what makes KDN true is that the 'ought' is grounded in an (actual or idealized) desire of the disagreeing parties to maintain or gain knowledge about the disputed issue. On another gloss, the relevant 'ought' is based on a ranking implicit in certain, say, social norms, thereby rendering the 'ought' in KDN a kind of deontic 'ought'. On a more robustly realist tack, one might think the 'ought' of KDN is tied to a valuational structure that is desire- and social norm-transcendent. We shall not attempt to adjudicate between these different views here, remaining silent for the most part on questions of meta-epistemology.

Fifth, and finally, conditions (i) and (ii) of KDN will have less bite to the extent that disagreement has the effect of automatically defeating knowledge or automatically defeating the knowledge-transferring capacity of trust. Now, no one thinks that *all* instances of disagreement have these defeat-effects. For, in many cases of disagreement—in particular, where only one of the two disagreeing parties is an expert, or where one party possesses more evidence than the other—it is obviously true that one can continue to know in the face of disagreement, and that one can come to know by trusting one's disagreeing interlocutor. For example, imagine that Tim believes no one is at home; he calls Ana on her mobile from his office, and expresses this belief. Ana disagrees—because she, in fact, is at home. Obviously Ana continues to know that someone (indeed, she) is at home, and Tim can come to know this himself by trusting Ana. While disagreement does not always destroy knowledge and the knowledge-transmitting power of trust, it is a live question whether it does so in a wide range of the more

interesting cases of disagreement. A vexed question, central to the disagreement debate, is whether knowledge is defeated in certain kinds of cases involving 'peers'.[15] Many favour a view on which knowledge is defeated in cases of peer disagreement. In general, the greater the number of cases in which disagreement defeats knowledge or the knowledge-transferring capacities of trust, the more cases of disagreement will be relegated to the auspices of (iii). That is, the more disagreement defeats knowledge or the knowledge-conferring power of trust, the more KDN will recommend suspending judgment. In the following discussion, we will assume for the most part that knowledge (and the knowledge-conferring power of trust) is left undefeated by disagreement. As a result, those who are sympathetic with defeatist views will find our descriptions of some cases of disagreement to be jarring or even incoherent. In due course, we shall discuss whether and to what extent various purported limitations of KDN can be overcome by taking the phenomenon of defeat more seriously.

4 The normative inadequacy of KDN

Many will already be unhappy with KDN. In particular, one might be worried about cases in which one is simply not in a position to know what specific action—being dogmatic, conciliatory, or suspending judgment—KDN recommends. For there are possible cases of disagreement in which one knows but fails to know that one knows,[16] and cases in which one doesn't know but isn't in a position to know that one doesn't know. (And even more obviously, cases where one's disagreeing interlocutor knows but one does not know that he knows—indeed any case where one's disagreeing interlocutor knows will fit this profile at the time of disagreement since if one knew one knew at t one wouldn't be disagreeing at t). In such cases, even if one knows that one ought to conform to KDN, one is not in a position to know what specific action one should undertake to achieve that conformity. We might say that in such cases, one is not in a position to know what KDN 'demands of one'. Imagine the following situation: Sally and Harry are disagreeing about whether p. In fact, p is true, and Sally knows this, but she isn't in a position to know that she knows this.[17] Harry (falsely) believes not-p, and (as very often happens), he is not in a position to know that he doesn't know this. Imagine further that Sally can maintain her knowledge that p by being dogmatic and that

Harry can come to know p by trusting Sally.[18] Since neither party is in a position to know the facts about knowledge relevant to KDN, neither party is in a position to know precisely what action KDN demands of him or her.[19] To be somewhat more precise, we might say that KDN is not *perfectly operationalizable*, where a norm N (of the form 'S ought to F in circumstances G') is *perfectly operationalizable* iff, whenever one knows N and is in G, one is in a position to engage in a piece of knowledgeable practical reasoning[20] of the form:

(1) I am in circumstances G
(2) I ought to F in G
(3) I can F by A-ing

where A is a basic (mental or physical) act type that one knows how to perform. As the case of Sally and Harry shows, KDN is not perfectly operationalizable. This is because one is not always in a position to know whether one knows, and not always in a position to know whether one's interlocutor knows. In other words, the relevant kinds of knowledge-related conditions are *non-transparent*, where a condition C is transparent just in case, whenever it obtains, one is in a position to know whether it obtains.[21] Since knowledge is non-transparent, KDN is not perfectly operationalizable.[22]

Prima facie, KDN's imperfect operationalizability is troubling, and seems to count against KDN. But how should we make sense of this intuition?

As a first pass, we might worry that its imperfect operationalizability makes KDN ineligible as *advice*. That is, while KDN might specify some ideal ranking of outcomes, it is not the kind of thing that can be offered to actual agents as a guide to action. For an ought-claim to be a suitable vehicle of advice (this line of thinking goes), it must use an ordering source whose application to candidate actions can be known. If the 'ought' relies on a mode of ranking such that one sometimes can't know how a candidate action is ranked, then that use of 'ought' is unsuitable for an advisory role.

---

[18] Of course there is an extended sense in which, insofar as Harry is in a position to trust Sally, he is in a position to come to know that his belief that not-p is not a case of knowledge. For once he comes to know p he can deduce that his former belief that not-p is not a case of knowledge. 'In a position to know' is normally used with a narrower ambit. And in any case Harry is not in a position to know whether KDN demands trust in advance of engaging in trust.

[19] There is also the case where Harry believes not-p and knows he doesn't know not-p but doesn't know whether Sally knows not-p. Here Harry is in a position to know that KDN recommends the cessation of belief but not in a position to know whether KDN recommends trust or instead suspension.

[20] By 'knowledgeable' practical reasoning we mean practical reasoning that involves knowing each of the premises involved.

[21] We borrow this terminology from Williamson (2000), ch. 4. Williamson defines a condition C as *luminous* just in case, whenever S is in C, she is in a position to know she is in C. Let us say that a condition C is *absence-luminous* just in case, whenever S is not in C, she is in a position to know she is not in C; and (following Williamson) that a condition C as *transparent* just in case it is both luminous and absence-luminous.

[22] Of course a little more than the transparency of knowledge would be needed for knowledge of the relevant counterfactuals. Still we take it the most obvious and central obstacle to the perfect operationalizability of KDN is the non-transparency of knowledge.

How far does this worry take us? First, note that it often depends on conversational context whether, in proffering a bit of advice, one presupposes operationalizability. Suppose you are advising Jane on the giving of an award, and you say: 'You ought to give the award to the person who just walked through the door.' Uttered in a typical context, this presupposes that Jane knows (or is at least in a position to know) who just walked through the door. But one could also reasonably advise Jane as follows: 'You ought to give the award to the most deserving person. I realise that it's often difficult to tell who the most deserving person is.' Here, one is recommending that the award be given to the most deserving person, but one by no means expects the recommendation to be operationalizable in the sense above. But so long as one does not falsely presuppose operationalizability, it is far from clear that there is anything *ipso facto* wrong about articulating an imperfectly operationalizable norm as advice. After all, there can be instances in which one can't incorporate a norm in knowledgeable practical reasoning but nonetheless has good evidence about what a norm recommends. Suppose Hanna gives you the advice: 'You ought to put out as many chairs as there are guests.' You have good evidence that there will be six guests, but you don't know this. Hanna's advice is hardly improper or useless, despite your not being able to incorporate it into knowledgable practical reasoning.

Indeed, even if offering a norm *as advice* presupposed a sort of operationalizability, this is at most a constraint on advice *at a context*, not in general. That is, just because there are cases in which KDN exhibits operationalizability-failures, this does not preclude it from *ever* being useful as advice; it will count as advice in those contexts, at least, when it is operationalizable. So while it is false that whenever we know, we know we know, it is perfectly plausible that there are plenty of disagreement cases in which we both know and know we know. In such cases, one might well know what KDN demands of one. (Of course one will never *know* KDN demands trust in a situation in which one's interlocutor knows p and one believes not-p and where such knowledge would be transmitted by trust—though insofar as one knows one doesn't know p one will be in a position to know that KDN entails that ought to stop believing p.)[23]

If the conditions relevant to KDN were transparent, then every (rational) attempt to conform to KDN would be successful. But since they are non-transparent, (rational) attempts to conform to KDN might fail. For this reason KDN can easily fail to be good advice because trying to follow it, or exhorting others to follow it, does not guarantee conformity with it. To what extent do these mismatches between trying and succeeding constitute a major deficiency of KDN?[24]

The first thing to say here is that it is no count against KDN in particular that *trying* to comply with it does not always succeed. For this fate will likely afflict any norm. This is

---

[23] Leaving aside recherché cases, if there be such, where one knows that one knows p and also believes not-p.

[24] Or we might think that this shows us that KDN is incomplete; to borrow some terminology from ethics, we might say that KDN is a 'criterion of right', but that what is (also) needed is a 'decision procedure' that tells agents how to *decide* what to do. Thanks to Andreas Mogensen.

very obvious in, say, the case of act utilitarianism, since trying to be a good utilitarian will (likely) result in less overall happiness than might have otherwise been enjoyed—say, by following the norms of commonsense morality. But it is also plausibly true with relatively more transparent norms. For example, suppose we had an ordering source that ranked actions according to how much one felt like doing them, and thus a norm that said that one ought to do what one most felt like doing. One is generally in a position to know what one feels like doing. But of course one can be mistaken about what one feels like doing, and one can be mistaken about *how much* one feels like doing something compared to something else. Insofar as the facts about what and how much one feels like doing things are non-transparent, then *trying to do* that which one most feels like doing will not always be a guaranteed way of doing that which in fact ranks highest (even if one always succeeds in doing what one tries to do). Since no plausible ordering source is transparent (we return to this theme shortly), no plausible norm is such that *trying* to do what one ought to do will guarantee in fact doing what one ought to do.[25]

Of course, one could still ask: what general advice should we give about responding to disagreement, given that we want to maximize conformity with KDN over some extended period of time? Probably, exhorting people to conform to KDN isn't the best way of doing this. Indeed, the answer to this question will turn on a vast number of open empirical questions about how people respond to disagreement and advice, relevant trade-offs between agents' storage capacity and the compliance-generation of particular bits of advice, and so on. This computation becomes even more vexed when we consider the effects that certain bits of advice will have in the long run or across a range of possible worlds. And it's worth remembering that the answer to this question might turn out to be very odd indeed. Given the oddities of human psychology, it might turn out that we ought to, say, advise people to, inter alia, drink a glass of water or take a deep breath when confronted by disagreement. In all likelihood, the correct answer to this question is not going to resemble the kind of norms that are put forward by epistemologists—whether it be KDN or any other candidate disagreement norm.

In sum, the fact that trying to comply with KDN will not guarantee compliance with KDN is itself of no especial concern, for it is plausibly true of any norm. If instead we simply want to know what general advice we should disseminate with regard to disagreement, then we have switched over to an empirical question that, while perhaps interesting, is well beyond the purview of the standard debate about disagreement.

Worries that turn on the advisory role of norms have not amounted to much. Here is a second (and we think, better) pass at what is worrying about KDN: it severs a natural tie between an agent's evaluative status—how praiseworthy or blameworthy that agent is—and facts about what the agent ought to do. The imperfect operationalizability of KDN generates cases in which one does what one ought to do (according to KDN) but is intuitively *blameworthy* for so doing, and conversely cases in which one does what one

---

[25] Note also that if there are no transparent conditions, then no decision procedure will be perfectly operationalizable either. (See fn. 24.)

ought not to do (according to KDN) but is intuitively *blameless* for so doing. Consider the following:

Clairvoyant Maud.[26] Maud is a clairvoyant, and uses her clairvoyance to come to know that the British prime minister is in New York, though she doesn't know that she knows this. Her friends, who are members of Parliament and therefore usually know the whereabouts of the prime minister, assure her that the prime minister is in fact at 10 Downing Street. Maud, moreover, doesn't even believe she is clairvoyant, as she has been exposed to plenty of evidence that suggests that clairvoyance is impossible. Nonetheless, Maud dismisses her friends and continues to believe that the prime minister is New York.

Let us stipulate that it is possible to gain knowledge through clairvoyance, and that although Maud's evidence that clairvoyance is impossible means that she isn't in a position to *know* that she knows that the prime minister is in New York, she nonetheless does know his location.[27] Then Maud, in being dogmatic, conforms to KDN; if she were instead to be conciliatory in the face of the disagreement, she would lose her knowledge that the prime minister is in New York. Nonetheless, it seems that Maud is doing something epistemically irresponsible by being dogmatic. We feel a strong intuitive pull towards the judgment that Maud is doing what she ought not do, for she is maintaining a belief even when she has overwhelming (albeit misleading) evidence that she isn't clairvoyant, and thus doesn't know the disputed proposition. We can't help thinking that Maud is playing with epistemic fire, exhibiting poor habits of mind that just happen, in this rare case, to serve her well. Thus, KDN allows for instances of what we might call 'blameworthy right-doing': that is, cases in which S intuitively does something blameworthy, though according to KDN she does what she ought to do.

   Cases of blameworthy right-doing can also be generated in instances where one trusts a disagreeing interlocutor whom one has (misleading) reason to believe doesn't know. For example, imagine that one of Maud's friends, John, despite his evidence that Maud is not clairvoyant and thus *doesn't* know that the prime minister is in New York, trusts Maud and comes to believe that he is. Here, John is doing what he ought to do—assuming that the knowledge-transmitting capacity of trust is not defeated by the disagreement—though it seems, intuitively, that he is blameworthy for doing so. In both of these instances of blameworthy right-doing, the agents conform to KDN while not being in a position to know that they are doing what they ought to do. Generally, these instances of blameworthy right-doing are instances where S conforms to KDN but where it is likely on S's evidence that she doesn't so conform. Let's call this *evidential* blameworthy right-doing.

---

[26] This case is adapted from Laurence BonJour, 'Externalist Theories of Empirical Knowledge', *Midwest Studies in Philosophy*, 5 (1980), 53-73. BonJour uses this case to argue against a simple reliabilist theory of knowledge, on which S knows *p* just in case *p* and S uses a reliable method to believe *p*. BonJour concludes from the Clairvoyant Maud case that having overwhelming but misleading evidence that one's method *isn't* reliable defeats knowledge. In response to BonJour's case, many externalists embrace a view on which justification/knowledge can be defeated by misleading evidence. We discuss this 'defeatist' view shortly.

[27] In other words, her knowledge that the prime minister is in New York is undefeated by the evidence against clairvoyance.

Arguably, there might also be instances of blameworthy right-doing where one *knows* one is conforming to KDN. For example, consider the following case:

Bridge Builder. Simon is an expert bridge engineer. He is constructing a bridge to span a large river, which thousands of commuters will cross each day. Simon has done the relevant calculations, and knows precisely how many struts are required to hold up the bridge. Simon's colleague, Arthur, a more junior but competent engineer, disagrees with Simon's assessment, saying that more struts are required.

Let us stipulate that Simon not only knows how many struts are required, but also knows that he knows this. Arthur, while almost always right himself, makes mistakes on a few more occasions, and Simon knows this. According to KDN, Simon should dismiss Arthur and be dogmatic about the number of struts required. Indeed, Simon is in a position to know that he should do this, since (*ex hypothesi*) he not only knows how many struts are required, but moreover knows that he knows. Nonetheless, if Simon were to simply dismiss Arthur, we would likely feel that this would be problematic. Why is this?

The problem isn't that for all Simon knows Arthur might be right, since Simon, *ex hypothesi*, knows Arthur is wrong. And the problem with dismissing Arthur can't be that for all Simon knows, it will turn out upon consulting with Arthur that Simon doesn't now know after all. If, as we are supposing in the present discussion, disagreement doesn't defeat knowledge, then knowing one knows renders an eventuality like Bridge Builder live.[28] What seems problematic about Simon's dismissal of Arthur is that Simon is instilling in himself a bad habit—that is, a habit of boldly going on even in the face of disagreement, a habit that might easily lead him to disastrous consequences. Our nervousness about Simon's dogmatism, we would like to suggest, turns on our recognition that if Simon were in a case where he in fact didn't know how many struts were required, the habit he is instilling in himself in the case where he *does* know might easily lead him to act similarly dogmatically, thus building an unsafe bridge and threatening the lives of thousands. Of course, if Simon were always in a position to know when he *didn't* know, there would be no such risk. That is, if Simon could always perfectly distinguish between cases in which he knows and doesn't know, the habit he is instilling in himself would be fine. But since there are not unlikely eventualities in which Simon isn't in a position to know that he doesn't know—again, because knowledge is non-transparent—the habit he is instilling in himself by dismissing Arthur is problematic. Human beings are not creatures for whom the absence of knowledge is generally luminous; as such, it is simply not possible for humans to be dogmatic in cases where they know and not also be dogmatic in cases where they falsely believe they know. That is, unless they are so selectively dogmatic in cases where they know they know that superficially similar situations do

---

[28] A defeat-friendly approach brings its own oddities. If Simon knows that he knows, and knows that disagreement will automatically destroy his knowledge, then Simon seems to have excellent reason not to consult with Arthur in the first place.

not arise.[29] We might call the kind of blameworthy right-doing displayed by Simon *habitual*.[30]

Conversely, KDN also results in cases of 'blameless wrongdoing', cases in which S intuitively does something that is *not* blameworthy, but where she does what (according to KDN) she ought not to do. For example, if Clairvoyant Maud were conciliatory in the face of the disagreement with her high-powered friends, or if Bridge Builder Simon were to doubt himself upon learning that Arthur disagreed with him, they would be doing something intuitively praiseworthy, or at least blameless. But according to KDN, they would be failing to do what they ought to do.

Let us call the fact that KDN generates instances of both blameworthy right-doing and, conversely, virtuous wrongdoing, the problem of *normative divergence*. There is a clear moral analogue to this problem. Take the following case:

Grenade. A soldier is holding a grenade that is about to detonate, and he must decide to throw it either to his left or to his right.

Let's assume that act consequentialism is the correct moral theory (or at least, more plausibly, that it is the correct moral theory with respect to Grenade). Then we might say that what the soldier ought to do is to conform to the following norm:

Consequentialist Norm (CN): If S is faced with the choice of doing only either A or B, S ought to do A if it would produce less harm than doing B, ought to do B if it would produce less harm than doing A, and is permitted to do either if A and B would produce equal harm.

Imagine that the soldier in Grenade has misleading evidence that more harm will be done if he throws the grenade to the right. If he throws the grenade to the right, then he does (according to CN) what he ought not to have done, for he performed the action that resulted in greater harm. Nonetheless, he is obviously not *blameworthy* for doing what he does. This is an instance of blameless wrongdoing. Now suppose instead the soldier throws the grenade to the left, because he wants to *maximize* the possible harm of his action. In fact, his action minimizes the actual harm done; nonetheless, we certainly don't want to say that his action was *praiseworthy*. As such, the claim that (as CN entails) the soldier ought to throw the grenade to the left does not supply the grounds for appropriate normative evaluation of the soldier's actions. Both KDN and CN, then, suffer from the problem of normative divergence. That is, both link 'ought' to an ordering source that implies that there is no straightforward tie between what agents ought to do and the evaluative status of their actions or their character.

This, we take it, is what is most deeply troubling about KDN: it fails to secure a naturally hoped-for tie between what agents ought to do and agents' evaluative status. To

---

[29] Suppose, for example, that while one often knows that one knows the result of an arithmetical sum, one sticks to one's guns only in cases where one has triple checked.

[30] Structurally similar issues arise when it comes to double-checking one's answer. Suppose Simon knows that he knows but is someone who would not know that he does not know in reasonably similar calculation situations. In that case, and especially when a lot is at stake, we might find it praiseworthy were he to suspend his belief only to reinstate it once he has double-checked his calculations.

accommodate this worry, one might try to modify KDN in one of two ways. One might think that, in addition to the 'ought' associated with KDN, we require a distinct kind of ranking of epistemic acts (yielding a distinct but associated use of 'ought'), a ranking that is directly tied not to the value of various *outcomes* but rather to the level of praise-worthiness/blameworthiness of the agent's *action*, considered as an attempt to achieve those outcomes. This modification would require two senses of 'ought' (an 'objective ought' and a 'subjective ought') as it applies to disagreement cases. Let us call this the *two-state solution*. However, it might be thought that bifurcation involved in the two-state solution creates an unacceptable rift between knowledge and epistemically virtuous conduct, and in particular, that this rift drains knowledge of its normative status. So, one might attempt to tinker with the conditions on knowledge such that the two phenomena—conforming to KDN and doing what is epistemically praiseworthy—line up. Note, after all, that in the most obvious cases of what we labelled 'blameworthy right-doing' and 'blameless wrongdoing', there was a mismatch between which act would in fact retain/produce knowledge, and which action was *likely* on the evidence to retain/produce knowledge. If, for example, we allowed Maud's evidence that she does not know to preclude her knowing—hence treating the original description of Clairvoyant Maud as incoherent—then those kind of putative cases of blameworthy right-doing would be ruled out. Let us call such a solution, on which epistemically blameworthy behaviour is incompatible with knowledge, the *defeatist solution*. Indeed, the desire to have an epistemology that captures our intuitions about epistemic blameworthiness seems to be a major motivation for standard defeatist views. Such views offer prima facie hope of making knowledge incompatible with non-virtuous epistemic conduct. By accepting such a view, we might hope to rescue KDN from the problem of normative divergence. We will discuss the defeatist solution in the next section, and the two-state solution in the one following.

5 Defeatism

How far does defeatism take us in overcoming normative divergence? Even if knowledge were incompatible with its being likely on one's evidence that one does not know, this will not suffice to collapse an outcome-driven 'ought' and a evaluation-driven 'ought'. For on any plausible account, the absence of knowledge is compatible with the likelihood of its presence. Suppose S does not know *p* but has plenty of evidence that she does, is confronted with disagreement, and sticks to her guns. In so doing, S is in violation of KDN, but intuitively blameless; this is an instance of blameless wrongdoing. At most, then, a defeatist solution will do away with certain instances of blameworthy right-doing (e.g. Clairvoyant Maud).

How well might a defeatist view deal with blameworthy right-doing? Recall that the phenomenon of blameworthy right-doing divided into two sorts: *evidential* and *habitual*. In an instance of *evidential* blameworthy right-doing (e.g. Clairvoyant Maud), S conforms to KDN though it is likely on her evidence that she is in violation of it.

In an instance of *habitual* blameworthy right-doing (e.g. Bridge Builder), S conforms to KDN, and it is likely on her evidence that she is conforming to it, but nonetheless she does something epistemically blameworthy by inculcating in herself a dangerously dogmatic habit.

Let us take *habitual* blameworthy right-doing first. In these cases, S knows *p* and knows that she knows *p*, but in conforming to KDN inculcates in herself a habit that makes it likely that she will be dogmatic in similar cases in which she does not in fact know. According to the defeatist solution, on which blameworthy epistemic behaviour is incompatible with knowledge, such cases are impossible. Is this a plausible response? To get a better grip on such cases, let us take a practical analogy. Suppose two tennis players are each in situations where they know they know they can't reach the ball in time. One player—call him Andi—gives up. Another, call him Raphael, chases the ball though (unsurprisingly) fails to get to it in time. Andi might deride Raphael as a pathetic figure, someone who gives chase while knowing the chase is futile. But we can see that because the absence of knowledge isn't luminous, Andi risks turning himself into a player who fails to chase the ball when he might, in fact, reach it in time. For cases will arise sooner or later where Raphael *believes* he knows he won't reach the ball, but in fact doesn't know this. In such cases, thanks to the habit he has inculcated in himself, Raphael will sometimes end up reaching the ball. On the other hand, Andi doesn't chase the ball both in cases where he knows that he won't reach it, *and* in those cases in which he falsely takes himself to know that he won't reach it. In this way, Andi—in failing to chase the ball even when he knows he won't reach it—inculcates in himself an undesirable habit. While we thus might all agree that there is something untoward about Andi's behaviour on the court, it seems very odd to think, as a defeatist view suggests, that his giving up in the good case costs him knowledge. Similarly, there is something untoward about dogmatism. But it is similarly odd to think that untoward dogmatism costs knowledge.

Consider a version of Bridge Builder in which Arthur does not disagree with Simon, but instead voices a much milder protest. Suppose Simon calculates the number of struts required and comes to know that he knows that twelve struts are needed. Suppose then Arthur expresses a little epistemic apprehension: 'Perhaps you should double check. It would be really awful if you made a mistake.' Simon dogmatically presses forward and brushes off Arthur's concerns. Here too we feel that Simon is getting into bad habits— but it would be rather far-fetched to suppose that Arthur's apprehension serves to defeat Simon's knowledge that he knows.[31] Plausibly, dangerous dogmatic habits do not generally cost one the ability to know.

---

[31] There is a contextualist model of what is going here with which we won't engage: Arthur's apprehension puts Simon in a context where the relation he now expresses by 'knows' is one in which he does not stand to the fact that twelve struts are required, even if he continues to stand in the relation that he previously expressed by 'knows' to that fact (and even, perhaps, continues to stand in the relation to the fact consisting of his standing in that relation to the fact).

Now it might be thought that there is a crucial difference between the case in which Arthur merely expresses apprehension and in which Arthur actually voices a contradictory opinion. For on some tempting probabilistic models, the fact of Arthur's disagreement makes it *no longer likely* from Simon's perspective that twelve struts are required, thereby costing Simon his original knowledge. How so? Suppose, as seems natural, that for S to know *p*, *p* must have a suitably high probability on S's evidence. And suppose that prior to hearing Arthur's opinion, Simon's epistemic probability in the fact that the bridge requires twelve struts was sufficiently high for Simon to know it. Assuming that Simon knows that Arthur has a very good track record, the epistemic probability (for Simon) that Arthur will agree with Simon is high. But it is easy to have the intuition that the probability of the fact that twelve struts are required, *conditional* on Arthur's disagreeing with Simon, is not high enough for knowledge.[32, 33] But note that models like this are of no use if one thinks of knowing *p* as sufficient for *p*'s being part of one's body of evidence. For insofar as Simon's total body of evidence *includes* the fact that twelve struts are required, he will hardly be able to conditionalize his way to a less than high probability in this fact. It is worth underscoring the oddness of leaving out the relevant piece of knowledge from Simon's total body of evidence. As we naturally think about the case, we take all sorts of other bits of knowledge that Simon has as sufficient for rendering various facts part of his evidence. If, for example, we think of Simon as knowing various facts about Arthur's track record, we are *ipso facto* inclined to count those facts as part of what Simon has to go on. Leaving out the bridge-building facts that he knows from his body of evidence thus might seem somewhat ad hoc. At the very least, probabilistic considerations need not force us to accept a defeatist view on which habitual blameworthy right-doing in disagreements such as this is an incoherent phenomenon. Indeed, from a perspective according to which knowing *p* is sufficient for *p*'s being part of one's evidence, such a defeatist view is implausible.

What of the phenomenon of *evidential* blameworthy right-doing? In such cases, S conforms to KDN although it is likely on her evidence that she isn't so conforming. According to a defeatist view on which blameworthy epistemic conduct is incompatible with knowledge, such cases cannot arise. To flesh out this kind of view, fans of defeat often argue that something like the following is true:

Evidence-Bridge Principle (EBP): If it is likely on S's evidence that S doesn't know *p*, then S doesn't know *p*.

---

[32] As many have noted, such models face difficulties in dealing with disagreements about mathematics and logic, since standard probabilistic models assign probability 1 to all mathematical and logical truths. This precludes disagreement having any potential to lower probabilities via conditionalization. The challenge of making good on some notion of non-idealized probabilities is not an easy one to meet. We return to this theme briefly in the next section.

[33] Note that even if Simon's track record is known to be a bit better than Arthur's, that will not help much (vis-à-vis a setting where Arthur is considered a peer). Suppose Simon's conditional probability of his being right conditional on a disagreement is 0.6, of Arthur being right 0.4. That would still, on this model, give disagreement a knowledge-destroying effect (assuming 0.6 is too low for knowledge). Thus, on such a model, epistemic peerhood is not what is crucial.

EBP will not be tempting for defenders of multi-premise closure concerning know-ledge. By multi-premise closure, if one knows each of a number of premises and deduces their conjunction, then one knows the conjunction. But suppose each conjunct is such that one does not know that one knows it and, indeed, that for each premise there is some non-negligible likelihood that one does not know it. Then it will be easy to flesh out the case such that it is likely on one's evidence that one does not know the conclu-sion. (Except in special cases where the risks of failing to know each conjunct are highly interdependent, the risks will add up to a large risk given enough conjuncts.)

Of more general interest here is Williamson's recent persuasive case against any such bridge principle, one that shows its incompatibility with natural ways of thinking about margins of error.[34] Imagine you are looking at a pointer on a dial. Given the distance you are from the dial, the particular light conditions, and so on, there exists some margin of error n such that that there is some strongest proposition p you are in a position to know of the form *the pointer is plus or minus n degrees from point x*, where x is the actual position of the pointer. If you were to believe, say, that *the pointer is plus or minus n-1 degrees from point x*, you would not in fact *know* this proposition. Suppose, on this particu-lar occasion, the strongest proposition you know about the position of the pointer is the proposition p, that the pointer is within range Q. That is, for all you know, the pointer is anywhere within the range Q, a range which has position x, the actual position, at its centre. Now, note that nearly all of the positions within Q preclude knowing p. If, say, the position of the pointer were closer to the edge of Q than point x, then one's margin for error would preclude knowing p. So it is very unlikely, relative to the propositions that you know (including p itself), that you know p.[35, 36, 37]

The general upshot of Williamson's argument, we take it, is the following. Defeatism can be helpfully thought of as a view on which knowledge is what we might call a 'minimally luminous' state. A minimally luminous state is one such that whenever one is in it, it is not the case that it's *unlikely* on one's evidence that one is in it. But Williamson's argument suggests that, given some plausible assumptions about margins of error, knowledge is not even minimally luminous.[38] If one clings onto EBP in the face of such cases it seems that one will be forced to deny that there is any such thing as the strongest proposition one knows, which in turn seems to generate a forced march to a sceptical

[34] Timothy Williamson, 'Improbable Knowing' in Trent Dougherty (ed.) *Evidentialism and Its Discontents* (Oxford: Oxford University Press, 2011), ch. 9.

[35] Note that the argument makes no particular assumptions about how much you know about the posi-tion of the pointer, beyond the anti-sceptical assumption that in such a case one knows something about the position of the pointer by visual inspection.

[36] We also note in passing that the models that Williamson describe wherein one knows even though it is likely on one's evidence that one does not know can easily be extended to describe cases where one knows even though one *knows* that it is likely on one's evidence that one does not know.

[37] As Williamson notes, the argument will need to be complicated a bit further to account for 'inexactness in our knowledge of the width of the margin for error' ('Improbable Knowing', 155). We shall not go through the argument that incorporates variable margins here.

[38] Williamson's argument, like his anti-luminosity argument, generalizes to all non-trivial conditions. We focus here on knowledge so as to intersect with the debate about defeat.

conclusion that one knows nothing about the relevant subject matter. So we are faced with a choice between abandoning evidence- or justification-bridge principles on one hand, and embracing radical scepticism on the other.[39] Assuming Williamson is right, then not only can one not eliminate the case of habitual blameworthy right-doing using defeat considerations; non-transparency means that one cannot eliminate the case of evidential blameworthy right-doing either.

We certainly do not take ourselves to have shown that epistemologists are misguided in positing the phenomenon of defeat.[40] But we do take ourselves to have made trouble for one of the plausible motivations for defeatist views—namely, that of trying to align knowledge with epistemic virtue. One obviously cannot eliminate the myriad cases of blameless wrongdoing using defeat considerations. But of more interest is that one cannot plausibly eliminate the paradigmatic cases of blameworthy right-doing, either.[41]

6 The two-state solution

There is certainly some intuitive pull to the thought that, in addition to an 'ought' governed by outcomes, we need an 'ought' that is tied to praiseworthy epistemic conduct—just as it is natural to draw an analogous distinction in the moral realm between what one *objectively* ought to do and what one *subjectively* ought to do. That said, the praise-connected 'ought' is rather more elusive than it initially seems. In this section we explain why. Again, our explanation will turn on considerations of non-transparency.

A natural first pass on the 'subjective' epistemic ought will mimic its standard analogue in the moral sphere: *one ought to do that which has greatest expected epistemic utility*. Suppose that there exists some fixed scale of epistemic utility in the disagreement situation that is KDN-inspired.[42] Then, the idea goes, the praiseworthiness of epistemic conduct is given by this scale in combination with the facts about the subject's particular epistemic situation. Such an 'ought' can in principle sit perfectly well alongside an 'ought' whose ordering source is generated by a ranking of outcomes. The introduction of this subjective 'ought' is thus not an objection to KDN; instead, it is supplementary structure designed to remedy those ways in which KDN is silent. One might also argue that this subjective 'ought' is truer to the context in which philosophers or the folk raise

---

[39] Similar arguments can be formulated against analogous bridge principles, for example, ones that rely on *safety* or *reliability* as the salient notion.

[40] For a general anti-defeat case see Maria Lasonen Aarnio, *Indefeasible Knowledge* (DPhil dissertation, University of Oxford, 2009) and 'Unreasonable Knowledge', *Philosophical Perspectives* 24, 2010. The present discussion is significantly influenced by those writings.

[41] These considerations also raise the question of whether continuing to believe when it's likely on one's evidence that one does not know is even intuitively blameworthy in general. After all, it is far from clear that we find the pointer-looker intuitively blameworthy for believing the strongest proposition that she knows.

[42] It can be at best inspired and not determined by KDN. To fix a scale of this sort will require far more than a ranking that, say, puts knowledge on top, withheld belief second, knowledge-less true belief third, and knowledge-less false belief fourth. A full valuational structure would both rank outcomes and provide valuational *distances* between them.

questions about how to deal with disagreement, and as such is a better answer to the question with which we began.

That said, the 'subjective' epistemic ought might fail to capture our intuitive rankings of epistemic praiseworthiness. Here are three reasons why. First, it is easy to imagine cases in which one is not in a position to know of the presence or absence of a piece of evidence: that is, cases in which one's *evidence* is non-transparent. This is especially obvious if one's evidence is that which one knows: if one knows *p* but is not in a position to know that one knows *p*, then one is not in a position to know that *p* is part of one's evidence. Conversely, if one *doesn't* know *p* but is in no position to know that one does not know *p*, then one is not in a position to know that *p* is not part of one's evidence. This possibility will again generate instances of blameworthy right-doing and blameless wrongdoing. Suppose, for example, that Sally justifiably takes herself to know certain false propositions about Harry's reliability; she has lots of misleading evidence that Harry is extremely reliable. Relative to everything that Sally justifiably takes herself to know, the course of action with maximum expected epistemic utility is for her to trust Harry, but relative to what Sally *in fact* knows, the course of action with maximum expected epistemic utility is dogmatism. Using the subjective 'ought'—that is, the 'ought' tied to expected epistemic utility—we can say that Sally ought to be dogmatic. However, intuitively, if Sally were instead to trust Harry, she would be blameless for so doing. Thus, an 'ought' tied to expected epistemic utility will not necessarily match our intuitions about praiseworthy and blameworthy epistemic conduct.

This result is unsurprising, of course, given a conception of evidence on which knowing *p* is necessary and sufficient for *p*'s being part of one's evidence. Knowledge is non-transparent, and non-transparency gives rise to instances of normative divergence. Thus, one might reasonably think that we are working here with the wrong conception of evidence, a conception that is unsuited to our intuitions about praiseworthy and blameworthy epistemic conduct. However, this problem plausibly cannot be avoided by switching to a different notion of evidence. For assuming that no non-trivial condition is such that either its absence or presence is luminous,[43] there will always be the possibility of a mismatch between one's evidence and what one is in a position to know about one's evidence.

A tempting thought here is that the subjective 'ought' is a measure of what is best *by one's own lights*. But that thought becomes less tempting once we realize that whatever our gloss on 'our lights', there will plausibly be cases in which agents are justifiably mistaken about their own lights. In that case the phenomenon that gave rise to blameworthy right-doing and blameless wrongdoing with respect to the 'objective' ought—namely,

---

[43] For a persuasive case against the luminosity of any non-trivial condition, see Williamson (*Knowledge and Its Limits*), ch. 4. One can also generate less nuanced (but perhaps equally convincing) arguments for this conclusion. For example, consider the condition of *being in pain*, which seems to be as good a candidate as any for a luminous condition. It seems highly plausible that with sufficient psychological priming, one may inevitably mistake a pain for an itch or a sensation of intense cold.

a mismatch between the facts pertinent to what one ought to do and what one takes those facts to be—re-emerges for the 'by one's lights' ought. In short, if we introduce an 'ought' tied to expected epistemic utility, then the phenomena of blameworthy right-doing and blameless wrongdoing will still arise relative to *that* 'ought', again because of the non-transparency of evidence.

A second potential source of limitation in the expected epistemic utility model concerns mismatches between the actual probabilistic connections and what one justifiably takes them to be. Suppose, for example, that one is deciding whether to trust someone about a certain proposition that is in fact a complex theorem of classical logic. If epistemic probabilities are standard, at least to the extent that all logical truths are assigned probability 1, then the facts of the disagreement will be probabilistically irrelevant. The proposition will have probability 1 relative to all facts, and the expected epistemic utility of trusting one's interlocutor will be calculated accordingly. It is obvious enough, then, that any such conception of probability will induce intuitively compelling cases of blameless wrongdoing and blameworthy right-doing. But it is not obvious that we can contrive a non-idealized notion of probability that will provide a more satisfying gauge of praiseworthiness and blameworthiness.[44] Note that even if the operative notion of probability were subjective probability, that will not avoid the general worry, since there is no reason to expect that subjective probabilities are themselves luminous. This is especially clear if subjective probabilities are a matter of one's dispositions over all bets, since there is no guarantee that one is in a position to know one's own dispositions. But even if one thinks of assigning a subjective probability to a proposition as more akin to, say, feeling cold than being disposed to bet, anti-luminosity arguments for feeling cold will still apply.

Third and finally, the expected epistemic utility conception also fails to take account of the habitual considerations advanced earlier. What those considerations seem to indicate is that there are cases in which one knows that a certain action has greatest expected epistemic utility (using a KDN-inspired scale), but in which one is, nonetheless, blameworthy for doing it and blameless for not doing it. *Ex hypothesi*, Bridge Builder Simon knows that he knows that twelve struts are required, and hence (on very plausible assumptions) will be in a position to know that sticking to his guns has maximum expected epistemic utility (at least assuming that knowing *p* is sufficient for *p* to count as

---

[44] Note that in general one source of instability in our practices of praising and blaming others is the extent to which we take badly formed beliefs as nevertheless providing an excuse for some course of action. Consider the case of Frank, who is deciding how to respond to a disagreement about whether-*p*. Frank's disagreeing interlocutor knows not-*p* and Frank's evidence provides strong support for the hypothesis that the interlocutor knows not-*p*. But Frank is in the grip of a bad argument whose conclusion is that the disagreement is sufficient to preclude his interlocutor's knowing. Should we say that Frank ought to ignore the bad argument and go ahead and trust the interlocutor? Or should we say that Frank ought not to trust the interlocutor given the (misguided) argument he finds persuasive? Similar issues arise when contriving a 'subjective' ought for people with misguided moral convictions. See Gideon Rosen, 'Skepticism About Moral Responsibility', *Philosophical Perspectives* 18 (2004): 295-313; see also Michael J. Zimmerman, 'Moral Responsibility and Ignorance', *Ethics* 107 (1997): 410-28.

evidence).[45] But as we have said, it is easy to get into a frame of mind where, despite these facts, we think of his sticking to his guns as a worse course of action than, say, shifting to agnosticism.

Note that habitual distinctions can even make a difference to expected epistemic utilities once we shift from a KDN-inspired scale to one that looks at the longer term. Even if one thinks of knowledge as the hallmark of epistemic success, one might grade an action such as sticking to one's guns by long-run consequences for knowledge rather than by the very short-term consequentialism encoded by KDN. Similarly, one could have a subjective 'ought' that was geared to longer-term consequences. Consider a situation where S knows that she knows, but also knows that by being dogmatic now she will make it likely that she will dogmatically cling to false beliefs in the future. Here, being dogmatic will straightforwardly have maximum expected utility for S according to a KDN-inspired scale, but may well not have maximum expected utility when a longer-term scale is in play.

What we have seen so far is that a two-state solution that uses KDN-inspired expected epistemic utility as the ground for the so-called subjective 'ought' does not do what it is designed to do, namely generate an 'ought' that aligns with our intuitive judgments about praise and blame. This is because non-transparency is inescapable: as we attempt to index our 'ought' to more stably accessible or 'subjective' conditions, we come up against possible cases in which agents fail to be in a position to know whether those conditions obtain—and thus fail to be in a position where they are blameworthy for failing to do what they ought to do, or vice versa.

7 Disagreement, despair, and beyond

We have suggested that those of us who hope for a general and intuitively satisfying answer to the question that is at the centre of the disagreement debate—namely, what we ought to do, epistemically speaking, when faced with disagreement—might be hoping in vain. There are deep structural reasons why such an answer has proven, and will continue to prove, elusive. Intuitively, we expect epistemic norms to be *normatively* satisfying: that is, we expect them to track our intuitions about blameworthy and praiseworthy epistemic conduct. An epistemic norm that ties what one ought to do to a non-transparent condition (e.g. knowledge) is an epistemic norm that will not satisfy this basic desideratum. To construct an epistemic norm that is normatively satisfying, then, we require an epistemic 'ought' that is tied to only transparent conditions; unfortunately, no such conditions plausibly exist. As such, the hope of finding a normatively satisfying answer to the disagreement question seems like a hope unlikely to be satisfied.

---

[45] We say 'on plausible assumptions' because, for example, one might know that one knows p in a disagreement situation and yet conceivably not know whether, if one sticks to one's guns one will suddenly start believing p for bad reasons that are not one's current basis for believing p.

Where does this leave us? One kind of reaction to all this is to despair of any cogent treatment of non-ideal cases. For example, we might say that expected epistemic utility works as a measure for epistemic praise and blame for creatures for whom evidence and probabilistic connections are completely transparent.[46] (Habitual wrongdoing can't arise for such creatures, since that phenomenon requires the non-luminosity of the absence of knowledge.) And we might contend that for non-ideal creatures there is no stable measure of epistemic praise and blame, and that associated 'ought' claims are not ultimately coherent. Since we are squarely non-ideal creatures, we think this is despair indeed.[47]

Another kind of response finds fault in the attempt to formalize praiseworthiness in the guise of KDN-inspired expected utility. One might hold that facts about one's evidence provide some reason to do this or that, but that facts about what one takes one's evidence to be, as well as facts about what habits a course of action inculcates, also provide reason to do this or that. And one might hope, moreover, that there exists, at least in many cases, a fact about what one has 'all things considered' reason to do. Less committedly, one might envisage an ordering source that directly ranks acts in terms of comparative praiseworthiness, perhaps without trying to give any sort of quasi-reductive account of what grounds these facts of relative praiseworthiness. At least one problem with these responses is that they fail to notice the ways that praise and blame reactions can be tied to various parameters that cannot plausibly be brought together into an 'all things considered' or 'on balance' judgment. Consider again Andi, who gives up chasing the ball in cases where he knows that he won't reach it. If we focus on the local scenario, Andi's actions seem unassailable; after all, he gives up when he knows it is futile. By contrast, Raphael's behaviour seems a little lamentable; after all, he keeps chasing when he knows he can't reach the ball. But if we shift to a more global outlook, we can see Andi's course of action as problematic and blameworthy, Raphael's as noble and praiseworthy. It is far from clear that one of these outlooks has any special authority over the other.

The preceding reflections point to what some will take to be the most promising option apart from despair: namely, to claim that even within the realm of the more 'subjective' ought, there is context-dependence according to the kinds of weightings of values that are operative at a context. There isn't a single ordering source associated with evaluating the subject 'by her own lights'. Rather, perhaps, there are a range of candidate-ordering sources to which one might semantically tie 'ought' claims, and which one might allow to dictate one's reactive attitudes. Note that such context-dependence will likely infect other normative terms like 'reasonable' and 'rational'. From this perspective, it is an illusion to think one has fixed one's subject matter by saying 'I am using the "ought" of rationality here', for there will simply be *no* privileged 'ought' of rationality. One pressing question for this approach is how it handles the question as to which of

---

[46] Thanks to David Chalmers for pushing this point.

[47] Moreoever, if we require that the evidence of one's interlocutor be completely transparent as well as one's own, that makes the ideal case particularly far removed from ordinary disagreement. And if knowledge suffices for evidence it precludes disagreements where one of the parties knows.

the various reactive attitudes—praise, blame, and so on—one ought to have. It's natural to think that the context in which one says 'Andi ought to give up' should also be a context in which it is appropriate to say 'People ought not to condemn Andi for giving up', and indeed is a context where one condemns the person who condemns Andi, and so on. But this kind of contextualism is one where 'ought' seems to lose much of its motivational power once the contextualism has been reflectively absorbed. If one overhears someone saying 'Andi ought to be praised for giving up', it is hard to be much moved by that when one is aware that at other contexts, some can speak truly (about the same action) by saying 'Andi ought to be condemned for giving up'.

Finally, others will be tempted by a more expressivist reaction: 'ought' claims express kinds of recommendations, though there are both different styles of recommendations (the bouletic style has a different flavour from the deontic style, and within the deontic style, some are made in an advisory spirit, some in an optative spirit), and different moods in which we make them (sometimes we are looking at the longer term, sometimes the shorter term; sometimes we are driven by certain wishes and concerns, sometimes by others). How much either contextualism or expressivism really buys us over outright despair we leave to the reader to judge.

Our goal was to suggest that a natural hope—that we might settle on an intuitively satisfying and principled answer to the disagreement question—might remain unsatisfied. This becomes clear, we think, when we take seriously the non-transparency of any condition that might plausibly play a role in an epistemic norm. Since knowledge is obviously non-transparent, norms like KDN fail to live up to our normative expectations in some fairly obvious ways. But alternatives, like the claim that *one ought to do what has greatest expected epistemic utility*, again fail to satisfy our evaluative intuitions. We don't mean this to be a serious recommendation of KDN as an answer to the disagreement question. Indeed, our intention has been to suggest that there seems to be no single privileged answer to the question 'What ought we to do, epistemically speaking, when faced with a disagreement?' This thought, bleak as it might be, easily leads to bleaker thoughts. We have not argued for the conclusion here, but it seems that non-transparency poses a more general problem for the ambition to formulate all sorts of epistemic norms.[48] If so, then it is not just a stable answer to the disagreement question that will remain elusive, but also stable answers to *all* questions concerning what we ought to do, epistemically speaking.

---

[48] For a general discussion of the implications of non-transparency for normative theorizing, see Amia Srinivasan's "What's in a Norm?" (in progress).