Epistemic Consequentialism and Epistemic Trade-Offs

Jeff Dunn
DePauw University
jeffreydunn@depauw.edu

January 26, 2016

1 Introduction

A trade-off is a scenario where you have the option of sacrificing some value or creating some disvalue in exchange for even more value. Such scenarios have long done work in ethics. For instance, Judith Jarvis Thomson (1985) famously asks us to consider the scenario where by pushing a man off a footbridge into the path of a runaway trolley I can save five workers who will otherwise be struck. This is a trade-off: sacrifice one life to save five. Thomas Nagel (1980) similarly asks us to consider a scenario where I must twist the arm of a young boy in order to persuade his mother to lend me the keys to her car so I can take my injured friend to the hospital. This is a trade-off: I cause the boy some pain in exchange for saving my friend's life. Trade-offs have often been used to cast doubt on consequentialist approaches in ethics. It's not hard to see why. The consequentialist maintains that the right action is the one that leads to outcomes with the most value; thus, consequentialists look like they have to say that one is required to push the man off the footbridge and twist the young child's arm. To many, these verdicts are counterintuitive.

In epistemology there can be trade-offs, too, though they haven't received as much attention as their counterparts in ethics. Mirroring ethics, such trade-offs are normally taken to be challenges to consequentialist approaches in epistemology. What is a consequentialist approach in epistemology? Roughly, the epistemic consequentialist identifies what has epistemic value and disvalue, and then seeks to explain or analyze the normative facts in epistemology (whether these are facts about epistemic rightness, or obligation, or justification, or rationality, etc. depends on the specific view at

hand) in terms of conduciveness to epistemic value.[1] One recent approach that exemplifies this approach is the "accuracy first" program, initiated by James Joyce (1998). According to this program, one argues in favor of epistemic norms by showing that every doxastic state that violates that norm is worse in terms of accuracy (in a specified sense) than are doxastic states that adhere to the norm. In particular, Joyce (1998) shows that for any degree of belief function that is probabilistically incoherent, there is a probabilistically coherent degree of belief function that is more accurate than the incoherent one at every possible world.[2] This is meant to be an accuracy-based reason for having probabilistically coherent credences. Underlying this style of argument seems to be the consequentialist idea that one ought to adopt belief states that are conducive to accuracy. But once one says this, it looks like one is required to accept epistemic trade-offs.

Critics of epistemic consequentialism have accordingly used such trade-offs to argue against the view. Suppose that accuracy of belief is the only epistemic value. There can be cases where if I form what I am certain is an inaccurate belief now, I thereby stand to garner much more accuracy later.[3] For instance, Fumerton (1995) asks us to consider a scenario where I am nearly certain God does not exist. However, in order to win a grant I must convince the funding body that I am a believer and the only way to do this is to genuinely believe. If I get the grant, I will garner many more true beliefs than I otherwise would. My evidence suggests that in adopting the belief that God exists I am almost certainly adopting a false belief. Should I take the trade-off? Just as in Thomson's trolley case or Nagel's arm-twisting case, it is natural to feel that I shouldn't take the trade-off; instead I epistemically ought to follow my evidence and believe that God doesn't exist.

More recently, Hilary Greaves (2013) has given similar kinds of objections specifically to the accuracy first program mentioned above. Consider

> **Garden Imps:** Suppose that there is one imp playing in the garden before me, who I clearly see. Five other imps are inside and will come out to play in the next moment with some prob-

[1] For more on specifying epistemic consequentialism, see Berker (2013a,b) and Dunn (n.d.).

[2] For other representative work in this program see Greaves & Wallace (2006), Gibbard (2008), Joyce (2009), Leitgeb & Pettigrew (2010a,b), and Pettigrew (2012, 2013a,b,c, 2015, 2016).

[3] As Jenkins (2007) has noted, we don't need two separate times like this to get trade-offs. Just imagine a case where I have only two options: believe the surely false P and also the surely true Q and R, or fail to believe all of P, Q, and R.

> ability. Let I_0 be the proposition that the imp in front of me is playing in the garden, and I_1, I_2, ..., I_5 be the propositions that each of five other imps come out to play. Suppose the probability that I_n $(1 \leq n \leq 5)$ is true is $1 - (0.5 \times c(I_0))$, where $c(\cdot)$ is my credence function. Thus, if I can get myself to set $c(I_0) = 0$, I can certainly get five perfectly accurate beliefs. If I follow my evidence and set $c(I_0) = 1$, then I cannot be as accurate about I_1 – I_5, since each is only 50% likely to be true.

Again, it might seem that one shouldn't ignore one's evidence in a case like this and set $c(I_0) = 0$. Nevertheless, the consequentialist looks like she must say that one should.[4]

In this paper, I'll attempt to respond to these kinds of trade-off objections to a certain form of epistemic consequentialism, in particular, the form of epistemic consequentialism we get from the accuracy first program. One kind of response is to argue that the accuracy first program doesn't actually require one to endorse objectionable trade-offs.[5] This will not be my tack. Instead, I will show how even if the accuracy firster *is* committed to unintuitive trade-offs, the unintuitiveness of such trade-offs can be explained away.

I'll do this by adapting two different lines of thought from two different thinkers. The first line of thought comes from Thomas Nagel. After introducing the arm-twisting case, he writes:

> ... things will be better, what happens will be better, if I twist the child's arm than if I do not. But I will have done something worse. If considerations of what I may do, and the correlative claims of my victim against me, can outweigh the substantial impersonal value of what will happen, that can only be because the perspective of the agent has an importance in practical reasoning that resists domination by a conception of the world as a place where good and bad things happen whose value is perspective-free. (pp. 180-1)

The idea here is that if anything can make it wrong to take a trade-off it must have something to do with the perspective of the agent herself on the world.

[4] Similar trade-off cases are presented by Littlejohn (2012) and Berker (2013a,b) as explicit challenges to epistemic consequentialism.

[5] This is the approach taken by Konek & Levinstein (forthcoming) and Joyce (forthcoming).

The second line of thought comes from William Clifford. Clifford, of course, maintains that it is morally wrong to hold beliefs that go against one's evidence. In the course of his argument for this famous claim, he writes:

> If a belief is not realized immediately in open deeds, it is stored up for the guidance of the future. It goes to make a part of that aggregate of beliefs which is the link between sensation and action at every moment of all our lives, and which is so organized and compacted together that no part of it can be isolated from the rest, but every new addition modifies the structure of the whole. No real belief, however trifling and fragmentary it may seem, is ever truly insignificant; it prepares us to receive more of its like, confirms those which resembled it before, and weakens others; and so gradually it lays a stealthy train in our inmost thoughts, which may someday explode into overt action, and leave its stamp upon our character for ever.

The idea here is that ignoring one's evidence is always a dangerous *moral* policy.

These two ideas, from Nagel and Clifford, can be put to use in responding to trade-off objections to epistemic consequentialism. These might seem surprising allies for an epistemic consequentialist. For Nagel suggests that it is wrong to take the trade-off and twist the child's arm, and Clifford claims that it is always wrong to ignore one's evidence—the very thing the epistemic consequentialist says one should sometimes do in order to accrue accuracy. But good allies they are nevertheless. For Clifford's worry is that there are harmful *moral* consequences of beliefs that go against the evidence. But it's consistent with this that there is nothing *epistemically* wrong with sometimes believing against the evidence. And something like this is what I will argue: the epistemic consequentialist can maintain that there is nothing epistemically wrong with adopting a belief against the evidence in order to accrue accuracy, but such beliefs do have negative *pragmatic* consequences, which explains our disapproval of such beliefs. But do such trade-off beliefs really have negative pragmatic consequences? Yes. Because of the way in which our beliefs form our perspective of the world and thus as a guide to action, there is a precise sense in which a trade-off belief is to be discouraged on pragmatic grounds. In particular, I'll show that by not adopting a trade-off belief, one avoids worst case expected loss. This is where Nagel's insight is adopted: the reason we might want to avoid trade-off beliefs is rooted in

the fact that an agent's beliefs form her perspective on the world, and this perspective has an importance in practical reasoning.

2 The Maximizing Agent and the Evidential Agent

In responding to the trade-off objection, it will be useful to introduce two fictional characters: the maximizing agent and the evidential agent. The evidential agent (I'll call her 'Evie') is perfectly successful in believing at each time in a way that is supported by her evidence at that time. We can think of this in the following way. Let the function p_E give us the probability that a proposition is true conditional on whatever total evidence, E, Evie has. Following Williamson (2002), we can call p_E an *evidential probability function*. I assume throughout that there is a unique value for p_E on any evidence E. If c_E is Evie's degree of belief function given evidence E, then we can put this succinctly by saying that $p_E(X) = c_E(X)$ for all propositions X that Evie takes a doxastic attitude towards.

The maximizing agent (I'll call him 'Max') is the kind of agent who accepts trade-offs. For instance, he will believe that God exists in order to get the grant, and he will set his credence to 0 for the proposition I_0 in Garden Imps. I'll suppose that Max forms beliefs in such a way that he always maximizes expected accuracy over the long term.[6] Thus, he adopts at time t the credence function that has the maximum expected accuracy for him in the future.

To spell this out in a bit more detail it is useful to make use of the notion of a *dependence hypothesis*. In general, if I am choosing between some set of options $a_1, a_2, \ldots, a_n$, a dependence hypothesis is a statement of the form: *if it were that* a_i, *then it would be that* w, where a_i is one of my options and w is a possible world. Of course, my evidence might not indicate that some

[6] Why expected accuracy rather than just accuracy? Greaves (2013) gives a nice answer: "As in the ethical domain, there is a subjective and an objective notion of epistemic 'ought'. Ethically, one objectively-ought to perform the action that will in fact have the best consequences; but since agents typically cannot tell for sure which act this is, we also recognise a subjective 'ought'. An agent subjectively-ought to perform (roughly) the action that seems most likely to lead to the best consequences, given the agent's beliefs at the time of action: more precisely, an agent subjectively-ought to perform that action that maximizes expected utility or value, where the expectation value is taken with respect to the agent's own credences. In the epistemic case: objectively, there is little to say beyond that agents ought to have credence 1 in all and only true propositions. Our project is to develop subjective epistemic consequentialism." At any rate, I'm not here defending the idea that one should maximize expected accuracy. My aim is to argue that someone who maintains this can deflect the trade-off objection.

particular option a_i will result in some particular world w being actual. In that case, we'll want to assign probabilities to dependence hypotheses. One way to do represent this is with imaged probability functions.[7] In particular, let $pr(w \parallel a_i)$ be a function that tell us the probability that w is actual on the counterfactual supposition that a_i is chosen. To work out the expected value of some option, a_i, we simply define a value function for worlds, $v(w)$ and sum up over all such worlds:

$$\sum_w pr(w \parallel a_i) \times v(w)$$

Now, we can tailor this basic picture to fit the demands of the sort of epistemic consequentialism that seems to underwrite the accuracy first program. Following Michael Caie (forthcoming), we can call this view *credal consequentialism*. According to that view, belief states are representable by credence functions, accuracy of credence functions (closeness to the truth) is the sole thing of epistemic value, and what agents epistemically ought to do is determined by maximizing expected epistemic value. For credal consequentialism, then, the relevant options are credence functions that can be adopted. Let a particular credence function be denoted as $\mathbf{c}$. Since it is accuracy that is meant to be maximized, our value function will be a function that tells us how accurate a credence function is at a world. Let us denote this with $A(\mathbf{c}, w)$, which is a measure of the accuracy of $\mathbf{c}$ in w.[8] So far, then, this gives us that the expected accuracy of a credence function $\mathbf{c}$ is:

$$\sum_w pr(w \parallel \mathbf{c}) \times A(\mathbf{c}, w)$$

But this cannot quite be the whole story. For an epistemic consequentialist like Max is not only looking at the accuracy of the credence functions he is currently choosing between. He also wants to consider how his current choices will effect the accuracy of the credence functions he will adopt in the future. To handle this complication, we introduce a bit more complexity. We first add superscripts to denote the time at which the credence function is held. So, for instance, if $\mathbf{c}$ is held at time 2, this would be denoted as $\mathbf{c}^2$.

[7] Insert footnote describing an imaged probability function.

[8] It is standard to use scoring functions as measures of accuracy and for historical reasons, these tend to actually measure *inaccuracy* rather than accuracy. Since we're focusing on maximizing expected accuracy, one would have to make slight transformations to the traditional scoring rules, but this does not in anyway have an effect on the arguments here.

Then we have the expected accuracy of adopting $\mathbf{c}$ at time 0 ($\mathbf{c}^0$) as:

$$\sum_t \sum_w \sum_j pr(w \wedge \mathbf{c}_j^t \parallel \mathbf{c}^0) \times A(\mathbf{c}_j^t, w)$$

Notice that the dependence hypothesis has changed slightly so that the outcome of adopting a credence function is divided up into facts about the world itself (w) and facts about the credence function adopted in that world at a time ($\mathbf{c}_j^t$).

One final point of clarification is important. So far I've been using an unspecified probability function, pr, as weights for the expectation. But there are at least two ways to understand this probability function. According to the first way of going, the weight for the expectation is just the agent's own credences. According to the second way of going, the weight for the expectation is the function that encodes what the agent's evidence supports, p_E. These lead to different views about what maximization of expected accuracy entails in cases where agents do not always adopt the credences best supported by their evidence.

On the first way of going we think of the expected accuracy of S adopting $\mathbf{c}^0$ as weighted by S's own credences about the relevant dependence hypotheses. On the second way of going we think of the expected accuracy of S adopting $\mathbf{c}^0$ as weighted by what the S's evidence supports with respect to these hypotheses.

If we go with the first option, then there is a worry that agents might not have credences defined over the relevant dependence hypotheses very often. For instance, do you have any views about the credence function that you are likely to hold tomorrow if you assign a 0.8 credence to it raining this evening? Probably not. These are not the kinds of propositions to which we usually assign credence. But if so, then we'll often fail to get any verdicts about the expected accuracy of adopting one credence function over another.[9]

If you take the second option, and weight the expectation with p_E, then this worry is less pressing. For while I might not have formed any credences about dependence hypotheses, it is somewhat more plausible that my evidence does support—at least in certain cases—certain values for such hypotheses. For this reason, I'll focus on the version of epistemic consequentialism that weights expectations by p_E rather than the agent's own credences.[10]

[9] Caie (forthcoming) raises an objection to credal consequentialism that turns on this fact.

[10] There is, however, a potential worry here. If the credal consequentialist is looking

3 Trade-Offs and Worst-Case Expected Loss

Max is an accuracy maximizer. At each time he adopts the credence function that his evidence suggests will garner him the most accuracy. Evie follows her evidence. At each time she adopts the credence function that matches what her evidence supports. Max and Evie could thus be in identical evidential situations and yet form different credences. When this happens, it is important to note that Max is in an odd position with respect to his own credences. The fact that he, say, has a high credence that God exists doesn't mean he has any reason in favor of the proposition that God exists. Evie, on the other hand, doesn't face this odd situation.

This difference between Max and Evie might matter since agents use beliefs as the basis for decision and action. A belief that is held purely for its instrumental epistemic effects is not necessarily a good belief on which to base decisions or actions. But, then again, by taking trade-offs Max *is* making it more likely that he is accurate overall than had he not taken the trade-off. And more accurate beliefs tend to be a better basis for decision than less accurate beliefs. So, who does better here: Max or Evie?

Here's one way to think about the situation. Suppose that at every moment, for every proposition Max and Evie take doxastic attitudes toward, they will be offered a bet. They can each take either side of the bet, and will do so based on which side their respective credences indicate is advantageous. Suppose, too, that each proposition is treated equally in the sense that the total money to be won or lost on each bet is the same. Since Max maximizes accuracy over the long haul, over the long haul Max will do better than Evie since he will more often choose the right side of the bets.

But suppose things are slightly different. Suppose that Max and Evie are no longer offered bets on every proposition at every time. Instead, the propositions on which they are offered bets are chosen at random as is the time at which the bets are offered. In this case, Max's policy is riskier than Evie's. Max could be offered a bet on a credence he has adopted purely for its instrumental epistemic benefit, a *trade-off belief*. In that case, we would expect him to lose money, and since there isn't a constant stream of

to maximize expected epistemic value where the expectation is weighted by what the agent's evidence supports, then it looks as though evidence is playing a special role from the very beginning. And this might make one doubt that the view is genuinely one that puts accuracy *first*. Note, however, that this worry applies as much to a version of credal consequentialism that weights the expectations with c_E as one that weights the expectations with p_E. As I see it, this is an objection to credal consequentialism distinct from the trade-off objection. Accordingly, I don't aim to address it here. For more on this kind of objection, see Meacham (forthcoming).

bets being offered, Max may not have a chance to make up this loss. So, a suitably risk-averse agent might prefer, on practical grounds, to be more like Evie than like Max.

A simple model can illustrate this. Suppose that there are three propositions over which Max and Evie have credences defined, A, B, and C. Suppose that a trade-off is offered.

Trade-Off: If one gives full credence to A, which is known to be certainly false, then one is guaranteed to be right about B and C.[11]

Max, of course, takes the trade-off; Evie does not. We don't know which bets on which propositions (if any) Max and Evie might be offered. But we are interested in the worst-case. So, we can ask: for each set of bets they might be offered, what is the worst they might do?

To answer this question, consider what I'll call an *unfriendly bet*. An unfriendly bet is a bet where if you have a credence in X, it has as small a potential payout and as large a potential loss as you find favorable (with fixed total money in play, here fixed at £1). For example, if you are such that $c(X) = 0.9$, then the unfriendly bet for you on X is 11p if X, -89p if $\neg X$. Similarly, if $c(X) = 0.2$, then the unfriendly bet for you on X is -79p if X, 21p if $\neg X$.[12] What's the use of the notion of an unfriendly bet? Well, we're ignorant about the bets Max and Evie will be offered in two ways: (1) we don't know on which propositions bets will be offered, and (2) we don't know what the odds will be of those bets. Focusing on unfriendly bets takes care of the second uncertainty by making sure we are focusing on the worst case for each of them.

We still don't know which combination of bets Max and Evie might be offered. But it turns out that no matter which combination they are offered, the expected value for Max's bets are always equal to or less than the expected value for Evie's bets.

First consider Max. Max takes the trade-off so $c(A) = 1$. The unfriendly bet for him is thus one where Max gets 1p if A and -99p if $\neg A$. Since A is certainly false, the expected value for Max for a bet on A is -99p. Since he took the trade-off, Max's credences in B and C are certain to be correct. That is, either $c(B) = 0$ and $\neg B$ or $c(B) = 1$ and B (and similarly for C).

[11] Letting $v(A)$ be the truth-value of A, this is just to say that $c(B) = v(B)$ and $c(C) = v(C)$.

[12] I focus on favorable bets, because those are ones you seem compelled to accept. Nothing, however, changes in the argument if we were to define an unfriendly bet as a bet that has as small a potential payout and as large a potential loss as you find *fair*.

From this it follows that the expected values of the unfriendly bets on B and C are each 1p.[13]

Now consider Evie. Evie is such that $c(A) = 0$, since her evidence makes it certainly false. The unfriendly bet for her is one where she gets -99p if A, and 1p if $\neg A$. But since A is certainly false, the expected value for her on this bet is 1p. We don't know the numerical value of Evie's credences in B and C, but we do know that $c(B) = p(B)$ and that $c(C) = p(C)$, since Evie always matches what the evidence supports. The unfriendly bet on B for Evie when $c(B) = n$ is thus one where:

If $n \geq 0.5$ she gets $(101 - 100n)$ if B, and $(1 - 100n)$ if $\neg B$.

If $n < 0.5$ she gets $(100n - 99)$ if B, and $(100n + 1)$ if $\neg B$.

So, when $n \geq 0.5$, the expected value of the bet on B is:

$$p(B) \times (101 - 100n) + p(\neg B) \times (1 - 100n)$$

But since $p(B) = n$, this simplifies to:

$$n(101 - 100n) + (1 - n)(1 - 100n)$$

After some algebra, this is is just 1. The result is the same if $n < 0.5$, and obviously the same thing holds when C replaces B. So, for Evie, no matter the proposition, the expected value of each unfriendly bet is 1p.

We can summarize this in the following table:

	A	B	C
Max:	-99p	1p	1p
Evie:	1p	1p	1p

The top column lists the propositions on which bets could be offered. The values in the outcomes are expected values. No matter which set of bets are offered, then, the expected value for Max is worse than or equal to that of Evie. Importantly, this table represents the expected values of *unfriendly bets*, so we are here focusing on worst-case expected value. Thus, a decision rule that instructs one to minimize the worst-case expected loss would instruct one to be like Evie rather than like Max.[14] Is such a risk-averse rule

[13] For instance, if $c(B) = 0$, then the unfriendly bet is one where Max gets -99p if B and 1p if $\neg B$. But when $c(B) = 0$, it is certain that B is false.

[14] Note that we don't have to think of this as an actual decision problem. Presumably no one ever has a moment where they get to decide to be like Evie or like Max. Still, thinking of the problem as a decision problem can illuminate the advantages or disadvantages one would accrue in being like Evie or like Max.

appropriate here? Note that this is a scenario of extreme uncertainty: we don't know which bets will be offered or even with what odds they will be offered. This is a prime case, then, where such risk averse rules often have appeal. Further, making decisions on the basis of our beliefs is very much like being offered bets at random times on random propositions. We don't know ahead of time which of our beliefs are going to be called into service for decision-making. What this shows, then, is that there is a practical advantage to forming beliefs as Evie does rather than as Max does because of the way that our beliefs form our perspective on the world for the purpose of decision and action.

In a moment, I will say more about how this responds to the trade-off objection. But for now, let me respond to a natural objection. In determining the worst-case expected loss for Max and Evie, I weighted the expectation with p_E, the function that encodes what Max's or Evie's evidence supports. But one might object that the weight should instead be given by a chance function that encodes the actual chance of the propositions being true or false. One response to this is to simply note that when it comes to worst-case expected loss, what matters is what Evie's and Max's evidence tells them they should expect to lose by adopting various credences. Thus, the weight should be p_E and not a chance function. But an alternative response is possible, too. Suppose that Evie is calibrated within m of the true chance of the propositions on which she bets. That is, when $ch(X) = x$, Evie's credence lies somewhere between $x+m$ and $x-m$. In these cases, the chance-weighted expected value of an unfriendly bet on X will be $-100m$. So, as long as the trade-off Max is offered results in a worse expected value than this, the result here still stands. More generally, let $\Delta_i(X) = |p_E(X) - ch(X)|$ represent for agent i, the difference between what i's evidence supports with respect to X and the true chance of X. Letting T refer to the trade-off belief for Max, so long as

$$\operatorname*{argmax}_T \Delta_{\text{Max}}(T) > \operatorname*{argmax}_X \Delta_{\text{Evie}}(X),$$

for all propositions X that Evie takes a doxastic attitude towards, then Max has a greater worst case expected loss than Evie. In the cases that are typically given, $\Delta_i(T)$ is very high. Garden Imps is a good example: the chance that an imp is in front of you is 1, but to get the epistemic reward the trade-off you must accept is to assign that proposition credence 0.

4 Error Theory

According to this picture, a risk-averse agent has a reason to not accept epistemic trade-offs. This reason is rooted in the special role that beliefs have for us: they are our perspective on the world and so our way of deciding what actions are best. The person who accepts epistemic trade-offs tampers with this perspective in a risky way that might have significant practical downsides. Here's an analogy. You know there will be some days in the future where you need immediate access to large sums of money. Despite this, you adopt a risky investment policy. Over the long-run it is likely you will maximize profits, but on certain days there can be big dips in the funds available. Thus, we see one way in which thinking about how our beliefs form our perspective on the world give us reason not to accept epistemic trade-offs.

That said, it is important to be clear about the kind of response to the trade-off objection that is being offered. The reason offered above to not accept epistemic trade-offs is not an *epistemic* reason. Instead, it is a *practical* reason. I've argued that we have practical reasons, but not epistemic reasons, to avoid trade-off beliefs.[15] But does this, then, do anything to disarm the trade-off objections with which we started? I think so.

First, consider the different kinds of decision rules adopted above with respect to the evaluation of credences. When it comes to epistemic evaluations of credences, the rule is to maximize expected accuracy. When it comes to pragmatic evaluations of credences, the decision rule is to minimize worst-case expected loss. The rule that instructs one to maximize expected

[15]It's worth noting that this is something of a surprising view. For some have given voice to the opposite sort of view: there are *epistemic* reasons to respect the evidence, but that it sometimes serves our practical interests to disregard it. Pascal's wager is a classic example of this sort. If Pascal is right, then it is in our practical interests to believe that God exists, but satisfying this practical interest might require us to go against the evidence and violate what seems to be our epistemic interests. More recently, Landes & Williamson (2013) have argued that there are practical reasons to respect the principle of indifference because in doing so one minimizes worst-case expected loss. Recently, however, Jon Williamson (ms) has argued that this is in tension with respecting one's evidence. To adopt the credence function recommended by the principle of indifference is sometimes to *go beyond* the evidence. And, on Williamson's view, our epistemic obligations are simply to respect the evidence. Thus, we have practical considerations in favor of ignoring evidence and epistemic considerations in favor of following it. In the case of epistemic trade-offs just discussed, I've argued we have the opposite of this. Practical and epistemic considerations *are* in tension, but I've argued that practical reasons push us toward respecting the evidence whereas epistemic reasons push us toward ignoring the evidence.

accuracy looks like it is the appropriate one when we are focused on epistemic evaluation. For the epistemic value of a belief state is a function of the accuracy, or expected accuracy, of *all* the beliefs in that state. It doesn't matter, when it comes to epistemic value, whether a belief is acted upon or not. The belief state of an agent is a representation of the world and given this, the accuracy of the whole thing matters to its epistemic assessment. Contrast this with the practical value of a belief state. Here it is not the practical value of *every* belief that matters, but rather only the beliefs that are acted upon. So, in cases of uncertainty about which belief will be needed as a basis for action, a risk-averse rule—like one that instructs you to minimize worst-case expected loss—is appropriate. It is not as if these decision rules are simply chosen because they happen to give the result the credal consequentialist is after. These rules are natural ways to evaluate beliefs from these two different perspectives: the epistemic and the practical.

Second, the values on which we are focusing here are both closely related to the central role of belief. On the one hand, we are looking at accuracy. Of course this is central to belief. On the other hand, we are looking at the practical value of using a belief as a basis for action. This is central to belief, too. Note, in particular, that the practical value of a belief here is not just the practical value of *holding* a belief. As Pascal's Wager makes clear, if someone will reward you for simply holding certain beliefs, then this can serve as a practical reason to hold the belief. If someone threatens to shoot me if I do not believe that the earth is flat, then the belief that the earth is flat would have significant practical value to me. But *this* value is not in any way central to the role of belief. In contrast to this, the practical value of a belief in the previous section is based on the bets a belief would lead the believer to accept. So we are focused specifically on what we might call the *guidance value* of a belief. And being a good guide for action is a central role that belief plays.

So, we have two ways of evaluating beliefs, one appropriate when we are thinking of beliefs in terms of their accuracy and another when thinking of them in terms of their guidance value. Both these ways of thinking about belief reflect a central role of belief. And when we evaluate beliefs in one way we get that we should take epistemic trade-offs; when we evaluate beliefs in the other way, we get that we should not take epistemic trade-offs.

Given this, we have the resources to respond to the trade-off objection to credal consequentialism. The response is this. The credal consequentialists are right that you epistemically should take the epistemic trade-offs. But it is also true that you practically should not. The reason that you practically should not accept trade-offs is because trade-off beliefs do a poor job of

playing the action-guiding role of belief. Given that this is such a central role of belief, it is not surprising that trade-off beliefs strike us as *bad beliefs*. They are, in a sense. It's just that they're not *epistemically* bad. So, the intuition that trade-off beliefs are bad is explained. They are *bad*; they are just not epistemically bad.

5 Objection

The best objection to the kind of argument I am making here is to find a case where (a) a trade-off belief is overwhelmingly unlikely to lead to any bad practical decisions and yet (b) it is still intuitively clear that the trade-off should not be accepted. A case meeting both these conditions would cast doubt on my claim that our intuitive verdicts against trade-off beliefs is a function of their practical, rather than epistemic, badness.

Here is a case that purports to meet both conditions. It is based on Greave's Garden Imps case. To ensure that condition (a) is met, we must add that the agent will not have to act on the basis of her credence that there is an imp before her now (I_0). This addition is important. For there are many things one might choose to do differently or not do at all conditional on an imp being in front of one rather than not. Suitably adjusted, this *Purified Imps* case certainly meets condition (a) and while it is unclear to *me* whether it meets condition (b), others have reported to me that they think it does.[16]

However, I think this objection can be answered. I have two main responses. First, when we compare Purified Imps to a more realistic but structurally analogous one, it is not clear whether condition (b) is met. Second, even if there are a small number of cases where the intuitive negative verdict about accepting trade-offs cannot be explained away by appealing to practical costs, there are other plausible explanations that can be given to take care of this small remainder of cases.

First, let's consider the realistic case. There is psychological evidence that, in the right sorts of circumstances[17], groups that are attempting to solve intellectual problems[18] reach a final answer that is accurate more often

[16]Thanks to the participants at the 2015 Bristol-Groningen Conference in Formal Epistemology for discussion on this point.

[17]Hastie (1986, pp. 151-2) identifies three characteristics that produce high levels of group performance: (1) the problem has a "eureka solution", a solution that may not be obvious initially but is demonstrable once discovered, (2) individual judgment accuracy is perturbed by unsystematic errors, and (3) group members possess different evidence.

[18]Examples of such problems include identifying the best candidate for a job, solving a

when the group members disagree with each other and defend their divergent views compared to when group members hold the same view. That is, groups that have disagreeing members who argue with each other and then reach consensus end up being more accurate in that final consensus than groups that all hold the same view, discuss the matter, and then reach consensus. Intriguingly, the benefits of diverse groups only accrue to those groups where members genuinely *believe* that different answers are correct; mere devil's advocacy—playing the role of a naysayer—does not have the same effect on the final accuracy of the group.[19]

Now, suppose Alice is one member of a two-member group, and she realizes the other member, who she regards as a peer, holds a different view as to the correct answer. Many think that in such a situation Alice's evidence supports withholding belief on the answer to the disputed question. This is because antecedent to the disagreement, Alice had no reason to discount the other group member's opinion and she has little reason to think that she is more reliable in this area than the other group member.[20] Grant that this plausible claim is true. Then, Alice faces a trade-off. Alice's evidence supports withholding belief in the answer to the disputed question and yet Alice can maximize her expected accuracy by maintaining her belief and arguing the matter with the other member of the group. By not being conciliatory, Alice can guarantee that the group members do not all hold the same view in the ensuing discussion (supposing the other member doesn't, improbably, simply switch to Alice's view upon registering the disagreement). To make things more concrete, suppose that the two members of the group work for an investment firm and are attempting to determine whether to hold some shares or sell them. They each know that no action will be taken with respect to the shares until after the meeting has adjourned. So there is no risk of acting on a belief not supported by the evidence, and thus no practical downside to taking the trade-off.

Notice that this case is structurally analogous to Purified Imps. There is no practical downside to ignoring one's evidence and in so-doing one maximizes one's overall accuracy. However, in this case, it seems to me that

logic puzzle, or deciding on the best investment strategy.

[19]Evidence for this includes Kuhn *et al.* (1997), Gigone & Hastie (1997), Moshman & Geil (1998), Schulz-Hardt *et al.* (2002), Perret-Clermont *et al.* (2004), Greitemeyer *et al.* (2006), and Schulz-Hardt *et al.* (2006). Strauss *et al.* (2011) and Mercier & Sperber (2011) each provide a summary of some of this experimental work.

[20]Many authors have defended something like this, for instance, Adam Elga (2007, 2010) who calls it the *equal weight view* and Christensen (2007, 2011) who calls it the *conciliatory view*.

condition (b) is not met. In fact, it seems to me that Alice should maintain her belief during the meeting in an attempt to reach the most accurate answer by the end of the meeting. At the very least it is not intuitively clear that Alice should follow her evidence and adopt a more neutral view during the meeting. So, while this case meets condition (a), it doesn't meet condition (b). But this, I think, casts doubt on whether Purified Imps meets condition (b), too. Is it really so unintuitive that one should take the trade-off in that case, too?

Some may still be unconvinced. Some may maintain that it is unintuitive to accept the trade-off in Purified Imps. This brings us to our second response. Recall that to get a case with the requisite structure, we needed to purify Garden Imps to ensure that condition (a) was met. But in most cases it is not: in most cases we have no guarantee that we will not have to act on the basis of some belief. Further, the very oddness of the case should give us pause. It is a case where it is stipulated that a belief that normally could be important for action and decision is rendered impotent. In addition, it is a case with an odd sort of dependence between my beliefs and the behavior of certain imps. Our intuitive judgments in such strange cases surely should be discounted, at least to some degree.[21] And finally, it is not implausible that our intuitive judgments about the permissibility of various beliefs are triggered by heuristics that we could expect to sometimes get things wrong, especially in odd fanciful cases. Consider how this works in ethics. We can set up cases where a particular action has very good consequences even though it is a type of action that usually has very bad consequences. We can understand why we might be disposed to judge actions by their type even if we think these judgments misfire in odd cases of the sort constructed. It is, after all, easier to identify whether an action inflicts suffering on someone than whether its total causal consequences are positive or negative. Even when it is explicitly stated that the consequences do outweigh the suffering inflicted in the act, it is not surprising that our intuitions continue to register that the action is wrong. I hypothesize something similar in the epistemic case. It is, after all, easier to determine whether someone's belief fits the evidence (especially in an easy case where the evidence in question is direct visual evidence in favor of a proposition) than whether the belief

[21] In a different context, Alistair Norcross (pp. 146-47) makes a similar move in rejecting Temkin's argument against the transitivity of 'better than'. Norcross claims that it might initially seem that a 32 million-year life with 2 years of torture is worse than a 32 million-year life with a mildly bad hangnail for the duration. However, he goes on to say that this is not a scenario in which we should trust our intuitions precisely because it is such an unfamiliar scenario.

has long-term positive epistemic consequences. So, even when it is explicitly stated that the consequences do outweigh going against the evidence, it is not surprising that our intuitions continue to register that the belief is wrong.

6 Conclusion

It may be useful to briefly recount the trade-off problem for credal consequentialism and the response given here. The problem is that credal consequentialism looks committed to saying that one should adopt certain credences that one suspects are inaccurate in scenarios where so-doing will allow one to accrue more accurate credences overall. But this, it is alleged, is unintuitive.

The main response in this paper has been to agree that in some cases the acceptance of such trade-offs does seem unintuitive. However, we can explain away these intuitions without saying that the trade-offs really are epistemically impermissible. The contention here is that because our beliefs form our perspective on the world, there is a practical downside to accepting epistemic trade-offs. More precisely, one minimizes worst-case expected loss by not accepting epistemic trade-offs. And further, because one of the central roles of belief is to guide us in our choice of actions, it is natural to see this practical downside as attributable to the belief itself. This, I claim, can explain many negative intuitions against accepting epistemic trade-offs.

However, there may be a remainder of cases where there is no such practical downside and yet some still feel the trade-offs to be intuitively objectionable. I've argued that such cases are rare and that in realistic cases it is not altogether clear that the trade-offs *are* intuitively objectionable. And further, I've hypothesized that there might be other ways to explain away intuitions against trade-offs in this small remainder of cases. The trade-off objection, then, is one to which the credal consequentialist can respond.

References

Berker, S. (2013a). Epistemic teleology and the separateness of propositions. *Philosophical Review*, *122*, 337–393.

Berker, S. (2013b). The rejection of epistemic consequentialism. *Philosophical Issues*, *23*(1), 363–387.

Caie, M. (forthcoming). A problem for credal consequentialism. In J. Dunn & K. Ahlstrom-Vij (Eds.), *Epistemic Consequentialism*, Oxford University Press.

Christensen, D. (2007). Epistemology of disagreement: the good news. *Philosophical Review*, *116*, 187–217.

Christensen, D. (2011). Disagreement, question-begging and epistemic self-criticism. *Philosophers' Imprint*, *11*(6), 1–22.

Dunn, J. (n.d.). Epistemic consequentialism. *Internet Encyclopedia of Philosophy*, `http://www.iep.utm.edu/`.

Elga, A. (2007). Reflection and disagreement. *Noûs*, *41*(3), 478–502.

Elga, A. (2010). How to disagree about how to disagree. In R. Feldman & T. Warfield (Eds.), *Disagreement*, Oxford University Press. 175–186.

Fumerton, R. (1995). *Metaepistemology and Skepticism*. Rowman & Littlefield.

Gibbard, A. (2008). Rational credence and the value of truth. In T. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Oxford University Press, vol. 2.

Gigone, D. & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin*, *121*(1), 149–167.

Greaves, H. (2013). Epistemic decision theory. *Mind*, *122*, 915–952.

Greaves, H. & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, *115*(459), 607–632.

Greitemeyer, T., Schulz-Hardt, S., Brodbeck, F. C., & Frey, D. (2006). Information sampling and group decision making: The effects of an advocacy decision procedure and task experience. *Journal of Experimental Psychology: Applied*, *12*(1), 31–42.

Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), *Information Pooling and Group Decision Making: Proceedings of the Second University of California, Irvine Conference on Political Economy*, JAI Press. 129–157.

Jenkins, C. S. (2007). Entitlement and rationality. *Synthese*, *157*, 25–45.

Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, *65*(4), 575–603.

Joyce, J. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of Belief*, Springer. 263–297.

Joyce, J. (forthcoming). Accuracy, self-accuracy, and epistemic consequentialism. In J. Dunn & K. Ahlstrom-Vij (Eds.), *Epistemic Consequentialism*, Oxford University Press.

Konek, J. & Levinstein, B. (forthcoming). The foundations of epistmemic decision theory. *Mind*. `http://jasonkonek.com/FEUT.pdf`.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction*, *15*(3), 287–315.

Landes, J. & Williamson, J. (2013). Objective bayesianism and the maximum entropy principle. *Entropy*, *15*(9), 3528–3591.

Leitgeb, H. & Pettigrew, R. (2010a). An objective justification of bayesianism i: Measuring inaccuracy. *Philosophy of Science*, *77*(2), 201–235.

Leitgeb, H. & Pettigrew, R. (2010b). An objective justification of bayesianism i: The consequences of minimizing inaccuracy. *Philosophy of Science*, *77*(2), 236–272.

Littlejohn, C. (2012). *Justification and the Truth Connection*. Cambridge University Press.

Meacham, C. (forthcoming). Can all-accuracy accounts justify evidential norms. In J. Dunn & K. Ahlstrom-Vij (Eds.), *Epistemic Consequentialism*, Oxford University Press.

Mercier, H. & Sperber, D. (2011). Why do humans reason? arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–111.

Moshman, D. & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Educational Psychology Papers and Publications*, 52.

Nagel, T. (1980). The limits of objectivity. *The Tanner Lectures on Human Values*, *1*, 75–139.

Perret-Clermont, A.-N., Carugati, F., & Oates, J. (2004). A socio-cognitive perspective on learning and cognitive development. In J. Oates & A. Grayson (Eds.), *Cognitive and language development in children*, Wiley-Blackwell, chap. 8.

Pettigrew, R. (2012). Accuracy, chance, and the principal principle. *Philosophical Review*, 241–275.

Pettigrew, R. (2013a). Accuracy and evidence. *Dialectica*, *67*, 579–596.

Pettigrew, R. (2013b). Epistemic utility and norms for credences. *Philosophy Compass*, *8*, 897–908.

Pettigrew, R. (2013c). A new epistemic utility argument for the principal principle. *Episteme*, 19–35.

Pettigrew, R. (2015). Accuracy and the belief-credence connection. *Philosophers' Imprint*, *15*(16), 1–20.

Pettigrew, R. (2016). Accuracy, risk, and the principle of indifference. *Philosophy and Phenomenological Research*, *92*(1), 35–59.

Schulz-Hardt, S., Brodbeck, F., Mojzisch, A., Kerschreiter, R., & Frey, D. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology*, *91*(6), 1080–1093.

Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes*, *88*(2), 563–586.

Strauss, S. G., Parker, A. M., & Bruce, J. B. (2011). The group matters: A review of processes and outcomes in intelligence analysis. *Group Dynamics: Theory, Research, and Practice*, *15*(2), 128–146.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415.

Williamson, J. (ms). Epistemic consequentialism and the principle of indifference.

Williamson, T. (2002). *Knowledge and its Limits*. Oxford University Press.