Permissible Epistemic Trade-offs*

Daniel J. Singer – singerd@phil.upenn.edu
Forthcoming in *Australasian Journal of Philosophy*.
Penultimate Draft (March 28, 2018). Please cite the published version.

Abstract

Recent rejections of epistemic consequentialism, like those from Firth, Jenkins, Berker, and Greaves, have argued that consequentialism is committed to objectionable trade-offs and suggest that consequentialism's propensity for trade-offs hints at a larger problem. Here I argue that sanctioning trade-offs isn't a fault of a theory of epistemic normativity because there are permissible epistemic trade-offs. I give examples of permissible epistemic trade-offs in pedagogy, in changes of worldview, and in indirect epistemic decisions. I also show that views that sanction trade-offs have an easier time than their rivals explaining both why we ought to be open-minded and how arguments with suppositions get their argumentative force. These considerations don't eliminate the consequentialist's burden to respond to the objectionable cases, but they do undermine the idea that no correct theory of epistemic normativity properly sanctions trade-offs.

Much of the contemporary debate about epistemic normativity mirrors an older debate between consequentialists and their opponents in ethics. In both subdisciplines, consequentialists sanction trade-offs that their opponents reject. Recently in epistemology, opponents of consequentialism have gained the upper hand by drawing on examples of objectionable trade-offs and further claiming that there are no epistemically permissible trade-offs, contra consequentialists' commitments. Here, I will argue that there are permissible epistemic trade-offs. I'll give three different kinds of examples, and I'll show that views that sanction trade-offs have an easier time than their rivals explaining some common epistemic phenomena. These considerations don't eliminate the consequentialist's burden to respond to the objectionable cases, but they do undermine the idea that no theory of epistemic normativity can properly sanction trade-offs.

1 Epistemic Consequentialism and the Trade-off Objections

Epistemic consequentialism is the view that "the epistemically right (e.g., the justified) is to be understood in terms of conduciveness to the epistemic good (e.g., true

*I'm very thankful to Selim Berker, Julia Driver, Kate Elgin, James Joyce, Jason Konek, Ram Neta, David Plunkett, Keshav Singh, Meredith Tamminga, and the audience at the Chapel Hill Normativity Workshop for helpful feedback on these ideas.

belief)" (Ahlstrom-Vij and Dunn 2014, p. 1). Like ethical consequentialism, epistemic consequentialism is a framework for more specific views: Epistemic consequentialists might disagree about what epistemic goodness is (popular options include true belief and knowledge, and less-popular ones are justified belief and understanding), what epistemic rightness is (e.g. justification, rationality, warrant, or what one epistemically ought to believe/do), and what the conduciveness relation is (e.g. maximizing over the long term, or indirectly promoting). Process reliabilism (Goldman 1979) is a form of consequentialism, since reliabilism takes a belief to be justified when it is the product of a process that reliably produces true beliefs (as such it is a form of *indirect* consequentialism). The simple view that what one ought to believe is what would maximize the number of true beliefs and minimize the number of false beliefs that they have over their lifetime is another consequentialist view.

According to trade-off objections to consequentialism, consequentialists incorrectly sanction actions or beliefs that sacrifice the good of some for the overall greater good. In ethics, a prominent example of this is the organ-harvesting case (Foot 1967), in which a consequentialist doctor is committed to trading off the life of one patient to save five others. In the epistemic realm, trade-off objections work analogously: objectors say that consequentialists wrongly sanction trading off the epistemic goodness of some of one's beliefs for the overall goodness of their doxastic state. Examples of trade-off objections to epistemic consequentialism can be found in Firth (1978), Fumerton (1995, p. 12), Jenkins (2007), Greaves (2013), Berker (2013a), and Berker (2013b).

The most straightforward trade-off objection to epistemic consequentialism comes from Jenkins (2007). Here is that case as formulated by Andow (2016):

TRUTH FAIRY Suppose you start with no reason to believe that p is true and no reason to believe that it is false. The Truth Fairy is a very powerful being, and she makes you the following credible offer: you accept p as true, and she will make your epistemic situation very, very good overall. She will arrange for you to have many, many true, justified, knowledgeable beliefs, and very, very few false, unjustified or unknowledgeable ones. However, she does not guarantee that your trust in p itself will have any particular epistemic status as a result of her actions.

In this case, consequentialists are committed to thinking that you should believe p (i.e. that it's justified, rational, or warranted), but of course, the natural intuition is that you should suspend belief.

Berker (2013b, p. 373) shows that consequentialists can't avoid trade-offs by limiting

the power of fairies or other external factors. There are trade-offs that occur completely inside an agent and where the belief that is sacrificed is a direct constitutive means to having the beliefs with epistemic value:

INTROSPECTIVE FELLOW Anytime Ignacio forms a belief, he also forms beliefs about that belief. So when Ignacio comes to believe P, e.g. that grass is green, he also comes to believe that he believes P, believe that he believes that he believes P, and onwards (although we can say that the process stops at some finite level). Ignacio is incredibly bad at forming the first-order beliefs, since he only forms those beliefs by reading tea leaves. But, Ignacio is incredibly good at introspection, so he is overwhelmingly likely to be right in his higher-order beliefs.

Here, consequentialists seem committed to saying that the agent's first-order beliefs are right even though they are formed by reading tea leaves. Why? It's because the beliefs are "a constitutive means of, at the same time, acquiring an immense number of true beliefs (and, for that matter, vast amounts of new knowledge, greatly increased coherence in [his] web of beliefs, much deeper understanding of [his] belief-forming habits, and so on)" (Berker 2013b, p. 373).

More sophisticated trade-off objections have also been formulated, including trade-off objections to formal consequentialist views that understand the rightness of having a credence in terms of its conduciveness to the epistemic good (see Greaves 2013). In response, defenders of consequentialism have tried to skirt the counterexamples by reconsidering their application to particular consequentialist views (e.g. Ahlstrom-Vij and Dunn (2014) and Goldman (2015)). Some opponents of consequentialism take their argument to extend past the particular counterexamples though. Berker, for example, says that his goal is "not to argue by counterexample" (2013a, p. 365). Greaves (2013, p. 918) seems to agree and puts the problem in terms of epistemic consequentialism's propensity to accept 'epistemic bribes.'

According to Berker (2013a), the more general problem with consequentialism that the trade-offs bring out is that consequentialism ignores the "separateness of propositions," a notion that is meant to be analogous to the "separateness of persons" issue for ethical consequentialism.[1] Berker (2013b) revises his view by arguing that the problem with consequentialism is actually that it tries to understand epistemic rightness in terms of promoting or conducing to some end. Berker's idea is that if epistemic rightness is about the promotion of outcomes, there is a directionality mismatch with standard intuitions: Whether an agent is justified or rational is usually

[1]For background on the structural worry about ethical consequentialism, see Brink (1993).

viewed as a question of whether she *responded to* or *respected* her evidence, and these are essentially backwards-looking notions. On the consequentialist picture, epistemic rightness has the wrong directionality: if consequentialism were right, epistemic rightness would be forward-looking, since it'd be about *bringing about* epistemic consequences. For that reason, the view can't account for the structure of epistemic notions like justification and rationality, Berker claims.

I'll argue that there are some permissible epistemic trade-offs, which does not show against the arguments by counterexample against consequentialism. But, it does show that trade-offs are not always impermissible, as Berker suggests. After that, I'll argue that the consequentialist can account for some epistemic phenomena in simple ways that rejectors of trade-offs cannot. That brings out some theoretical virtues of accepting trade-offs. First though, let's consider what a permissible trade-off might look like.

2 *The Shape of Epistemic Trade-offs*

We can learn something about the shape of permissible epistemic trade-offs by looking at the objections to consequentialism a little more closely. The first thing to notice is that when we consider most the cases, we consider them from the first-person perspective, i.e. we imagine ourselves as the subject in the case. For example, TRUTH FAIRY is written in the first person, so it's natural to ask oneself "Would it be the case that *I* should have this not-apparently-true belief for the sake of having other epistemically good beliefs, were *I* in that scenario?" The objections to consequentialism from Firth (1978) and Fumerton (1995, p. 12) have this character as well. Let's call cases like this *first-personal-deliberative cases*.

First-personal-deliberative cases are particularly helpful to the opponent of consequentialism. As Shah (2003) discussed, first-personal doxastic deliberation exhibits the phenomenon of *transparency*, that "when asking oneself *whether to believe that p*, [one must] immediately recognize that this question is settled by, and only by, answering the question *whether p is true*" (2003, p. 447).[2] Given transparency, in first-personal-deliberative cases, the question of whether to believe p will be settled by whether p appears true, which is a question settled by our evidence (and maybe prior beliefs) about p, not the epistemic goodness potentially produced by holding that belief. So transparency means that from the first-person perspective, questions of

[2]Whereas Shah (2003) and Shah and Velleman (2005) take it that transparency holds as a matter of conceptual necessity, Steglich-Petersen (2006) denies that transparency obtains *as a matter of conceptual necessity*, as does Chappell (2005). All of the authors accept that the phenomenon is prevalent among real agents though, even if it's not conceptually necessary.

epistemic rightness will appear to be settled only by evidence and prior mental states, making it a backwards-looking notion, which is exactly the result the opponent of consequentialism wants.[3]

Because of that, in the discussion below, none of the permissible epistemic trade-offs I offer will be *straightforward* first-personal-deliberative cases. If there were such cases, they would have to involve an agent who doesn't exhibit transparency, or they would have to be cases where the only epistemic consequences in play are ones that can be understood in terms of the belief itself (and no other beliefs) being true. Since neither sort of case is common or intuitive, it won't be broadly convincing to use those to argue for my point here.

I'll offer three kinds of permissible epistemic trade-offs, none of which is straightforwardly first-personal-deliberative. In the first two kinds of cases, we'll be asking ourselves what is permissible for others to believe. In the third kind of case, we do consider our own beliefs in the first person, but we do that indirectly. For that reason, it too is not a case where transparency is involved.

To simplify the discussion here, I'll use 'what one ought to believe' or 'what one should think' as generic epistemic rightness terms. The reader should substitute in their preferred term in those places. I'll also sometimes assume that either true belief or knowledge is the unique epistemic good, since much of the extant literature also discusses the view this way, but nothing I say below hangs on that assumption.

3 Trade-offs in Pedogogy

The first kind of permissible epistemic trade-off I'll discuss involves pedagogy. Suppose that for the past few months, your 7-year-old daughter Samantha has been showing interest in spacecraft design. Today, after school, she was asking a lot of questions about how we know where spacecraft will go and how fast they'll get there. You want to help Samantha learn about this, so you decide to teach her some physics. You could teach her about simple Newtonian mechanics, which would allow her to get many close-to-true (but false) beliefs now, or you could teach her the General Theory of Relativity (and non-Euclidean geometry), most of which would go over her head.

In this case, I take it that the natural thing to do is to tell Samantha what she ought to think, rather than what is true. You might say something like, "You should

[3]Of course, one might think that epistemic rightness comes apart from questions of what to believe in the first person, perhaps because there is more than one kind of epistemic rightness. Even if there are different senses of epistemic rightness though, the trade-off objections seem to be playing off the one that's present in these deliberations.

think about this is in terms of Newtonian mechanics" and follow that with instruction about forces, lines that are in flat geometries, etc. The idea is that teaching her about spacecraft movement with the simpler (but false) theory first will enable her to later come to have true beliefs about how spacecraft really move (either by approximation to the truth or by serving as a stepping stone to learning about Relativity).

What you say to Samantha in this case strikes me as true; she ought to think about it in Newtonian mechanical terms.[4] But, were Samantha to think about it that way, she would be coming to have false beliefs now as a way of trading off for epistemically valuable beliefs in the future. So, this is a case of a permissible epistemic trade-off.

You might think that it's not literally true that Samantha ought to think about it in Newtonian terms. Rather, she ought only to use Newtonian mechanics as a rough guide to the truth or as a heuristic, while keeping in mind that it's strictly-speaking false, you might say. This strikes me as an odd approach to pedagogy though. In humans, beliefs (but not working assumptions, mere guides, or heuristics) play a particularly efficient role in guiding deliberation and inquiry, both of which impact future belief-formation. Because of our limited memories and computational power, in some cases, real humans ought to have false beliefs when those beliefs will help us get a better grasp on the rest of the world. In our everyday lives, the attitude that we take towards 24 hours being the amount of time it takes Earth to rotate about its axis (something most of us know is strictly-speaking false) is no different from the attitude we take towards grass being green. It's only in contexts like this one, where it's made salient that one of them is known to be false, that we're inclined to say that the attitude we have towards the first is not belief.

It's also important to notice that, in this case, Samantha has no reason to suspect that you might be misleading her. Of course, *we* know that she would be forming false beliefs by following your advice, but we also know that teaching children about Newtonian mechanics is the first step to getting them to understand relativistic physics later on. In order to get there, Samantha should form the false beliefs first. If that's right, the best interpretation of what's going on is that this is a permissible epistemic trade-off: Samantha ought to think about spacecraft movement in Newtonian mechanical terms, and in doing so, she permissibly sacrifices the epistemic goodness of the beliefs she has now for the goodness of her future beliefs. It's easy to generalize this case to other cases of pedagogy (including self-pedagogy), so let's turn to the second kind of permissible trade-off.

[4] Certainly, it is at least permissible for her, which is all that is really needed.

4 Trade-offs in Changes of Worldview

In some cases, we should have a belief in part because of its connections to other beliefs. For example, that a belief explains other things an agent believes counts as a mark in its favor.[5] Coherentists (e.g. BonJour 1988 and Davidson 1986) think that all belief rightness is this way, since they take the rightness of a belief to be determined by how the belief coheres with others. Of course, less extreme versions of this view, where coherence is seen as some kind of epistemic virtue, are extremely common. Let's call the rightness of a belief *outward-looking* when it depends on the belief's connections to the holder's other beliefs.

The second kind of permissible epistemic trade-off involves the goodness of beliefs that have outward-looking rightness being traded-off in changes of worldview, i.e. big picture, systematic changes in what we believe. To simplify the discussion in this section, I'll assume the goodness of a belief for an agent is partially a function of how well supported the belief is by the agent's lights. This isn't straightforwardly compatible with consequentialist views that take true belief to be the unique epistemic good, but I will note how the case can be modified for those views.

First, notice that not just any case of outward-looking rightness involves a trade-off: If Sandy knows some large class of phenomena that needs explaining and learns that P explains them, it could be that P is the right thing for Sandy to believe. It would be right for Sandy to believe P because of how that belief would relate to the other things Sandy believes (so the rightness is outward-looking), but there is no trade-off. For it to be a trade-off, the goodness of the belief in question would have to be sacrificed for the overall epistemic goodness of Sandy's doxastic state. But in this case, Sandy's belief that P is supported by abduction on the things he knows. So coming to believe P is a win-win situation, not a trade-off.

In contrast, consider Eli's case: When she was young, Eli was a member of an extreme religious sect that held that Earth was roughly 6,000 years old. A couple years ago, Eli went away to college and took a course on evolution. At that time, Eli rejected evolutionary theory. Eli admitted that evolutionary theory would be a simple and highly explanatory theory were it true, but she thought that the world wasn't old enough for evolution to have worked. Now, a year after finishing the course, Eli is reconsidering her stance on the age of Earth. When Eli thinks about her evidence that bears directly on the question of Earth's age, she takes her evidence to overall support

[5] Of course, this is an instance of a broader kind of view that takes theoretic virtues such as explanatory power, simplicity, and fecundity to be reasons for belief.

thinking that the earth is young. But, she recognizes that were she to believe that the earth is older, she would come to have many new reasons to form beliefs about the evolutionary origin of species, many of which she would then regard as knowledge. In this situation, I take it that it is permissible for Eli to start believing that Earth is older, and this case, unlike Sandy's, is a case of a permissible epistemic trade-off.

To see why, contrast Eli with her twin Teli. Teli was also brought up being given evidence and arguments that Earth is roughly 6,000 years old. But Teli, unlike Eli, never left the religious community. Because of that, Teli isn't familiar with the explanatory mechanisms of evolutionary theory, and she has never been put in a position where she might consider revising her belief about the age of the Earth. In this case, unlike Eli's, it isn't permissible for Teli to revise her belief about the Earth's age. What makes the difference? It's that, for Eli (but not Teli), revising her belief enables her to better the state of many of her other beliefs by tying them into a common explanatory theory. Eli can avail herself of the evolutionary explanations of the visible differences, similarities, and changes of living species. Teli is not in a position to reap similar benefits, which is why a change in her belief is impermissible. So what makes it permissible for Eli to revise her belief is the overall increase in epistemic goodness of her doxastic state that it enables.

You might think that Eli's case is like Sandy's. Just like Sandy, after taking the evolution class, Eli knows a lot of things that stand in need of explanation. So, isn't what makes it permissible for her to revise her belief that the revised belief best explains the new things she knows, not that it leads to greater overall epistemic goodness in her epistemic state? Don't all of Eli's beliefs (including the belief about the age of the Earth) end up better off? If so, then there is no sacrifice of the goodness of the one belief for the overall benefit, and hence, no trade-off.

In response, first notice that Eli's case is importantly different from Sandy's in that, in Eli's case, the changed belief is not the one that does the explaining. Sandy simply adopts the belief that is already abductively supported by the other things he believes. Eli would only be able to adopt the beliefs that are abductively supported by her other beliefs, if she could stop believing that the Earth is too young for evolution to have happened. But, when Eli reflects on the age of Earth, she is faced with evidence that overall supports thinking that Earth is young. Eli is stuck in what we might think of as an epistemic bottleneck:[6] she would be in a much better epistemic state overall if she could just get past this one sticking point about Earth being too young for evolution to

[6]Thanks to an anonymous reviewer for putting it in these terms.

have worked. Sandy is not stuck in a similar bottleneck.

Let's consider how the details of this case might be further fleshed out. Suppose the case involves a progressive change in Eli's beliefs: she first comes to believe that the Earth is old, and then later she comes to believe the evolutionary facts. On this precisification, Eli sacrifices the goodness of her belief about the Earth's age for the goodness of her other (future) beliefs. In the time after Eli adopts the belief about Earth's age but before she works through the implications, that belief isn't well-supported for her — it conflicts with the things she believes from her upbringing.[7] It is only after she goes through the process of reasoning through the implications of her new belief and connecting it to beliefs about evolutionary explanations that the belief about Earth's age comes to cohere with her other beliefs. So, in this way of precisifying the case, she sacrifices the goodness of her belief at the earlier time for the goodness of her overall epistemic state later.

More generally, in permissible changes in worldview that happen progressively like this one, although the vast majority of our beliefs end up better supported than they were previously, some beliefs may end up being less well supported. In cases where what makes a belief permissible to hold in the new worldview is its connections to other aspects of the new view, the agent may have little or no other reason to accept it. In those cases, the amount of support for that particular belief may decrease despite an increase in the overall goodness of the full picture warranting the change. When that happens, the epistemic status of those less-well-supported beliefs is sacrificed for the greater overall epistemic good. These cases are the second kind of permissible epistemic trade-offs.

If you're worried that a progressive adoption of a new worldview like this is always impermissible, consider a different way of fleshing out the details of the case: Suppose Eli just adopts the whole swath of new beliefs in a single step. Instead of first changing her belief about the age of Earth, she works through the potential implications of adopting that belief, and then she adopts that new combined doxastic state in a single change. Of course, real belief change rarely works this way: when we change our beliefs in major ways, it takes a while for us to realize the implications of those changes. Notice though that if we think of the case this way, Eli doesn't *sacrifice* the goodness of her belief about Earth, since there is no time where that belief

[7] If epistemic goodness is true belief, then a different case where Eli comes to believe something false that would then lead her to believe many truths and give up the false belief will be better. We could imagine, for example, that Eli starts by believing that the world is only a little older than 100,000 years, just old enough to explain the appearances of low level organisms. She then reasons from there that it's more likely that Earth is much older still. It still seems permissible for her to take that first step.

becomes less well supported. So this is not a trade-off case in a strict sense. But, this precisification of the case would still show against the trade-off-based objections to consequentialism. That's because in this case, revising her belief about Earth's age is permissible *in virtue of* it leading to the betterment of the overall epistemic goodness of her full set of beliefs. The difference between Eli and Teli, who could not avail herself of evolutionary explanations by changing her belief, is that Teli cannot better her overall belief state by revising her belief. What makes it permissible for Eli to change her belief is that changing her belief about the age of Earth enables a new and dramatically better systematic approach to understanding the natural world. So on this precisification, despite the case not being a trade-off in the strict sense, it still demonstrates that epistemic rightness can be dependent on epistemic goodness, which is the feature of theories that licenses trade-offs.

5 Indirect First-Personal Trade-offs

The last example of a permissible epistemic trade-off involves an indirect doxastic decision. Here's the case: You've just passed your qualifying exams as a graduate student in chemistry. You know that project α is ingenious and would give us new foundational insights into molecular bonds and shape the research field for generations. Project α is difficult though, and you know there would be many skeptics until the project takes root, which may take a half-century. So, if you pursue project α you expect to be less professionally successful and less happy than you'd like. The alternative is project β. The lasting impact of project β would be minimal, but it would produce some new knowledge that project α would not produce. In particular you would come to know whether Q. Knowing whether Q would advance our understanding of how a certain protein folds, but it would not generate much other knowledge. Project β is in a very hot area of chemistry right now though. So, if you pursue project β, you'll be a lot more professionally successful and more happy. If you pursue α, you will suspend belief about whether Q. If you pursue β, you will suspend belief about the numerous foundational insights project α would have created.

Is it permissible to choose β? Regardless of whether pursuing β is permissible all-things-considered (something I'll remain agnostic about), if you do pursue it, there is a sense in which you would be criticizable. What we'd criticize you for, were we your future colleagues for example, is the harm you would have caused to yourself and the field in epistemic opportunity cost: "Think of how far the field would have advanced were you to have pursued that other project!" we might say. The criticism can't be a

practical one though, since the practical aspects of the case count in favor of β. The criticism is about the loss of knowledge (or true belief or understanding). So, what the appropriateness of that criticism shows is that, *epistemically speaking*, what you should do is pursue project α, even if that's not what you should do all-things-considered (again, a claim I'll remain agnostic about).

Recall that choosing α produces a much better overall epistemic outcome than β, but you'll come to know whether Q iff you pursue β. So, given that you should pursue α (epistemically speaking), this is a case where you should trade-off your future knowledge of whether Q for the epistemic gains of the other project, and hence, it is a case of a permissible epistemic trade-off.

If you're not convinced this is a trade-off case, compare this case to paradigmatic ethical trade-off cases. Take Thomson's (1976) trolley case: there is an out-of-control trolley heading towards five people strapped to the tracks, and you have the opportunity to divert the trolley towards just one person. Taking the trade-off would be to divert the trolley. Doing so is trading the life of the one to save the lives of the five. The analogy is straightforward: Professionally, you're heading right for project β, but you have the opportunity to "epistemically divert" yourself towards project α, which is significantly epistemically better. But, pursuing α means that you won't come to know whether Q. So, diverting your professional trajectory would be sacrificing your potential knowledge of whether Q for the better epistemic outcome of project α.

The natural worry here is that the trade-off isn't *epistemic*. The choice is about what research project to work on, and it's common to think that epistemic questions are only questions about what to believe. Since we're focused on a choice of a research project (which is a 'doing' rather than a 'believing'), doesn't that show that the issue is practical and not epistemic?

I don't think it does. Despite the common assumption among epistemologists that we cannot, we do seem to evaluate some 'doings' as either epistemically permissible or epistemically impermissible. For example, consider someone who follows Pascal's advice to "act[] as if [one] believed" to "cure [one] of unbelief" in God (1941, Sec. III). A common critique of Pascal's suggestion is that, regardless of whether so acting is practically right, there is something epistemically amiss about it. That is an epistemic evaluation of an action. Consider also people who search out more evidence about a hard question or run statistical tests on data to learn more about the data's structure. These actions seem epistemically evaluable too: You could be epistemically criticized for biasedly seeking out evidence or praised for running the right statistical tests.

Sometimes, we deliberate about actions that will knowingly have substantial

impact on our future doxastic states, like by affecting whether we'll have enough evidence to support our beliefs about the data or whether we'll know whether Q. These kinds of decisions are about what to believe just as much as they're decisions about what to do. But, unlike the cases of doxastic deliberation that Shah (2003) considers, in these cases, your deliberation doesn't directly result in you having a belief. What makes these cases *indirect* doxastic deliberation is that you choose features of your doxastic state by choosing an action that will give rise to those features. These decisions are epistemically indirect, but they're still decisions about what to believe, so we can evaluate them along the epistemic dimension (in addition to the practical and moral dimensions).

Does it matter that in direct doxastic deliberation one is choosing the content of one's beliefs rather than just choosing whether to have a belief on a given topic or to have more evidence for one's beliefs? I don't think it does. That's because there can be cases of indirect doxastic deliberation where the deliberation is about particular contents. Pascal's case is an example: Pascal's suggestion is that one choose to act in a certain way in order to come to believe a particular content, namely that God exists. We can also modify the case above to make it one where you're also choosing the content of your beliefs by choosing a research project. Suppose, for example, that folks who work in the area of project β all accept a methodological claim about how to study protein folding. As of yet, you do not accept that claim, but you're certain that you'll come to believe it if (and only if) you choose project β, perhaps because you have sufficient inductive evidence from seeing others go down that path. Then by choosing project β, you would be knowingly (but indirectly) choosing the content of your future belief. So, if you think we can only epistemically evaluate deliberation about specific contents, this modified example still contains a trade-off.

Another natural worry here is that the reasoning overgenerates. Almost every decision we make and action we perform has an impact on what we believe. If I order a chicken burrito rather than a veggie one for lunch today, that will influence whether I believe I am eating meat later. So doesn't that mean that, if we accept the reasoning above, we must think that all of those actions are epistemically evaluable?

My personal inclination is to bite the bullet and think that all actions are epistemically evaluable.[8] One need not accept anything nearly that radical in order to accept my argument here though. For the purposes of this argument, one only need accept that *some* actions are epistemically evaluable (and that among them are cases

[8] I've given an separate argument for this in [citation removed].

like the trade-off I described above). I hope to have made that intuitively plausible by pointing to cases where epistemic evaluations of actions seem ordinary, like the cases of collecting new evidence, analyzing data using statistics, and getting oneself to believe by self-indoctrination. It's open to the reader to think that it's only a very small class of actions that are epistemically evaluable. Perhaps it's ones where it is (or should be) obvious or salient to the agent that the decision or action will have significant and particular doxastic impacts. In cases like that, you make an indirect doxastic choice — you choose between options that in part include changes to your doxastic state. And, in some such cases, I've argued, there are permissible epistemic trade-offs.

6 Some Benefits of Accepting Trade-offs

Where does this leave the discussion of epistemic consequentialism? The existence of permissible trade-offs shows that epistemic consequentialism gets something right about the structure of epistemic normativity, even if it gets similar cases wrong, as Berker and others have argued. In this section, I'll show that epistemic consequentialism's acceptance of trade-offs counts in its favor in another way. Countenancing trade-offs uniquely allows the theory to provide natural explanations of some common epistemic practices.

First consider what I call "positive arguments by supposition." Here is an example: You're having lunch with your colleague from cognitive science. Unlike your colleague, you doubt the existence of cognitive states, and she is trying to convince you they exist. Your colleague starts by asking you to suppose that there are such states: "Let's see what follows if there are cognitive states," your colleague might say. She then goes on to show you many theoretical benefits of accepting that supposition. "It would allow us to account for the apparent non-reducibility of truths about mental states to truths about neuronal states," she might say. If she offers you enough benefits of accepting the supposition and you don't take the supposition to be too wild, you might accept the original supposition. That is a positive argument by supposition.

I take this kind of argument to be a dime a dozen in academic discussions. How do these arguments work though? If you accept epistemic consequentialism, you have a natural explanation of what's going on: the arguments show you the epistemic goods that come from accepting the supposed proposition. Since rightness is explained in terms of goodness on this account, making salient the epistemic goods that come from accepting the supposition can convince you that it's right to accept the proposition. If

you reject the permissibility of epistemic trade-offs, you can't accept this simple story about how these arguments work, since you'd have to deny that learning about the epistemic goods produced by holding a belief can make it right.

In response, rejectors of trade-offs might claim that by offering you benefits of accepting the supposition, your colleague is showing that your evidence already supports the supposed proposition. But this move doesn't explain the difference in argumentative force between positive arguments by supposition and straightforward arguments from evidence. For one, arguments by supposition can sometimes work, not because the hearer already accepts entailments of the supposed proposition (as she must do were it just an argument from the hearer's evidence), but because the hearer finds the full package of views plausible despite formerly rejecting individual pieces of that package. (The reader is encouraged to notice the similarity between this kind of case and the cases of changes of worldview above.) Epistemic consequentialists have a straightforward explanation of that phenomenon: the supposed proposition is right to believe in virtue of its role in generating the whole package of beliefs. If you reject trade-offs, you can't have a similarly straightforward account of those arguments.[9]

Something else that consequentialists (but not rejectors of trade-offs) can easily explain is the common view that something like modesty, humility, or open-mindedness is epistemically good. This assumption plays different roles (and goes under different headings) in different literatures. In the literature on peer disagreement, for example, it's the main motivator for conciliatory views (see Christensen (2013)). In the formal epistemology literature, the view is proposed as a constraint on rational credence known as "regularity," which requires agent to not assign extremal probabilities (0 or 1) to contingent propositions (see Lewis 1980 and Hájek 2012, sec. 3.3.4).

If we grant that something like this is epistemically virtuous, what might explain it? It's not enough to say just that we might be wrong if we have too strong beliefs on a topic. After all, we might also be right. A more plausible explanation of the requirement to be modest can be seen in Lewis's (1980, p. 268) discussion of regularity for credences:

[9]Notice also that the consequentialist has a neat story about why we have arguments by supposition: These arguments are tools that allow us to avoid the transparency phenomenon while coming to believe something because of the benefits that holding that belief has. Compare how a moral consequentialist might use a tool to kill a dying animal. Most of us could not help an animal that is dying a slow gruesome death by killing it with our hands. But, killing it might be the right thing to do, so to overcome our inability to do it manually, we might use a tool. Arguments by supposition play a similar role in our mental economy, the consequentialist should say: they allow us to overcome our deliberative inability to accept a belief on the basis of the doxastic benefits that it brings about.

> [Regularity] is required as a condition of reasonableness: one who started out with an irregular credence function (and who then learned from experience by conditionalizing) would stubbornly refuse to believe some propositions no matter what the evidence in their favor.

Skyrms (1980, p. 74) gives a similar argument. The idea behind these arguments is that being doxastically stubborn might make us miss out on good beliefs we could have in the future, whereas being open-minded allows us to have those beliefs. These "arguments from stubbornness" are forward-looking: the rightness of having a certain doxastic feature comes from its epistemic benefits down the line. Naturally, consequentialists can easily endorse these arguments. Those who reject all trade-offs cannot. If one accepts that the prospect of future epistemic goods can impact the rightness of an attitude now, one will be forced to countenance trade-offs that exploit that forward-lookingness.

What these two examples show is that epistemic consequentialism's propensity for trade-offs isn't just a quirk to be explained away. Countenancing trade-offs is theoretically useful in understanding our epistemic practices. Views that reject the possibility of permissible epistemic trade-offs cannot offer similarly straightforward and natural explanations of these phenomena.

7 Conclusion

As I mentioned in the introduction, some opponents of consequentialism have suggested that admitting of any trade-offs is a problem for a theory of epistemic normativity. Berker (2013b) thinks that trade-offs bring out that consequentialism gets the directionality of epistemic normativity wrong, and Greaves (2013, p. 950) thinks that trade-offs amount to 'epistemic bribe[s].' The existence and theoretical usefulness of permissible trade-offs shows against these universal rejections of trade-offs.

Previous responses to Berker have tried to show that particular consequentialist views, like process reliabilism, don't countenance the trade-offs that Berker says they do. Both Ahlstrom-Vij and Dunn (2014) and Goldman (2015), for example, argue that on the process reliabilist picture, justification does have the backward-looking structure that Berker thinks is characteristic of epistemically normative notions, despite reliabilism also having some forward-looking elements. If those responses to Berker are right though, process reliabilism won't be able to account for the permissible epistemic trade-offs that I brought out here. So they avoid Berker's objections only to be subject to new ones.

Epistemic normativity seems to have both backwards-looking and the forward-looking elements. Developing a systematic account of that normativity is beyond the scope of the argument here. The goal here was only to show the need for such projects by showing that there are permissible epistemic trade-offs and that accepting such trade-offs can do useful epistemological work.

References

Ahlstrom-Vij, K. and Dunn, J. (2014). A Defence of Epistemic Consequentialism. *Philosophical Quarterly*, 64(257):541–551.

Andow, J. (2016). Do Non-Philosophers Think Epistemic Consequentialism is Counterintuitive? *Synthese*, pages 1–13.

Berker, S. (2013a). Epistemic Teleology and the Separateness of Propositions. *Philosophical Review*, 122(3):337–393.

Berker, S. (2013b). The Rejection of Epistemic Consequentialism. *Philosophical Issues*, 23(1):363–387.

BonJour, L. (1988). *The Structure of Empirical Knowledge*. Harvard University Press.

Brink, D. (1993). The Separateness of Persons, Distributive Norms, and Moral Theory. In Frey, R. G. and Morris, C., editors, *Value, Welfare, and Morality*, pages 252–289. Cambridge University Press.

Chappell, R. Y. (2005). Inquiry and Deliberation. Blog Post on *Philosophy, et cetera*. http://www.philosophyetc.net/2005/08/inquiry-and-deliberation.html.

Christensen, D. (2013). Epistemic Modesty Defended. In Christensen, D. and Lackey, J., editors, *The Epistemology of Disagreement: New Essays*, page 77. Oxford University Press.

Davidson, D. (1986). A Coherence Theory of Truth and Knowledge. In *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*, pages 307–319. Basil Blackwell.

Firth, R. (1978). The Schneck Lectures. In *Delivered at the University of Arizona in 1978*.

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 5:5–15.

Fumerton, R. A. (1995). *Metaepistemology and skepticism*. Rowman & Littlefield Publishers.

Goldman, A. (1979). What is Justified Belief? In Pappas, G., editor, *Justification and Knowledge*. Dodrecht D. Reidel.

Goldman, A. I. (2015). Reliabilism, Veritism, and Epistemic Consequentialism. *Episteme*, 12(2):131–143.

Greaves, H. (2013). Epistemic Decision Theory. *Mind*, 122(488):915–952.

Hájek, A. (2012). Interpretations of Probability. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2012 edition.

Jenkins, C. S. (2007). Entitlement and Rationality. *Synthese*, 157(1):25–45.

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In *Studies in Inductive Logic and Probability*, volume 2, pages 83–132. University of California Press, Berkeley. 83-113.

Pascal, B. (1941). *Pensées*. New York: Modern Library.

Shah, N. (2003). How Truth Governs Belief. *Philosophical Review*, 112(4):447–482.

Shah, N. and Velleman, J. D. (2005). Doxastic Deliberation. *The Philosophical Review*, 114(4):497.

Skyrms, B. (1980). *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. Yale University Press.

Steglich-Petersen, A. (2006). No Norm Needed: On the Aim of Belief. *Philosophical Quarterly*, 56(225):499–516.

Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, 59(2):204–217.