Epistemic Utility Theory and the Aim of Belief

Jennifer Carr — University of California, San Diego

It's widely accepted that rational belief aims at truth.[1] Objectively correct belief is true belief. How should rational believers pursue the aim of truth?

Epistemic utility theorists have argued that the framework of decision theory can explain what it means to aim at truth. By pairing decision theory with a distinctively epistemic form of value—gradational accuracy, degrees of proximity to the truth—we can provide arguments for various epistemic norms in terms of truth.

These arguments generally use either the notion of *expected accuracy* or the notion of *accuracy dominance*. For example: it's been argued that the reason why we should have probabilistically coherent degrees of belief is that it's the only way to avoid accuracy domination (Joyce, 1998, 2009), and the reason why we should update by conditionalizing on our evidence is that doing so uniquely maximizes expected accuracy (Greaves and Wallace, 2006; Leitgeb and Pettigrew, 2010b).

Caie (2013) and Greaves (2013) show that results depend on notions of dominance and expected utility that are different in important respects from the versions of dominance and expected utility used in standard (practical) decision theory. If we use the more familiar forms of expected utility and dominance, we can't justify the epistemic norms that epistemic utility theory had hoped to justify. Indeed, the prescriptions of epistemic utility theory conflict with these norms.

I argue that the things epistemic utility theorists often call "expected accuracy" and "accuracy dominance" can't really be the expected accuracy or accuracy dominance of epistemic states in any conventional sense. It's not clear what they are; so far we don't have a good philosophical interpretation of these pieces of math. Without a philosophical interpretation, they are ill-equipped to do the epistemological work they were meant to do. For example, just telling us that conditionalization maximizes some particular quantity—whatever it is—doesn't explain why we should conditionalize on our evidence.

In short, those of us who are attracted to the project of epistemic utility theory face a dilemma. We must choose between old and new versions of rules like dominance avoidance and expected utility maximization. Call the familiar, decision-theoretic versions of these rules **consequentialist rules** and the new versions that appear in epistemic utility theory **nonconsequentialist rules**.[2] If we choose the consequentialist rules, then we can vindicate the idea that rational belief has the aim of accuracy—but at the cost of giving up rules of probabilistic coherence, conditionalization, and other attractive epistemic norms. On the other hand, if we choose the nonconsequentialist rules, we can avoid predicting rational violations of probabilistic coherence, conditionalization, and so on—but at the cost

[1] See e.g. (Velleman, 2000), (Wedgwood, 2002), (Shah, 2003), (Gibbard, 2003, 2005).

[2] I use "consequentialist" in a quite thin sense: that the value of an option depends entirely on what the world is like if the relevant agent takes that option.

of giving up the idea that epistemic rationality is a matter of aiming at the truth. This means giving up our explanation for why we should obey these rules.

Faced with this dilemma, I argue that we should choose the second horn. But this means that before epistemic utility theory can claim to justify epistemic norms like conditionalization or probabilism, it needs to explain and justify the nonconsequentialist rules that it treats as fundamental. Otherwise, these "rules" are just so much uninterpreted formalism.

The plan for the paper is as follows: in section 1, I'll briefly introduce epistemic utility theory and its motivations. Section 2 introduces the central example of the paper. The kinds of cases I'm interested in are cases where the truth of a proposition at least partly depends on what degrees of belief ("credences") a rational agent adopts. I'll argue that with traditional decision-theoretic measures of expected utility, the rule *maximize expected epistemic utility* (understood as gradational accuracy) actually conflicts with conditionalization.

How, then, were epistemic utility theorists able to argue that conditionalization always uniquely maximizes expected accuracy? In section 3, I argue that the quantity that epistemic utility theorists have called "expected accuracy" is different in important ways from the consequentialist notion of expected utility that we see in decision theory. I call this new measure "nonconsequentialist expected accuracy" or "NEA." I argue that closely related problems afflict accuracy "dominance" arguments: traditional (consequentialist) accuracy dominance avoidance generates conflicts with coherence norms like conditionalization and probabilism. But what epistemic utility theorists have called the "dominance" relation is very different from the consequentialist dominance relation in practical decision theory. Call this new relation "nonconsequentialist dominance" or "N-dominance." The prohibition of N-dominated credences doesn't lead to conflicts with epistemic norms like conditionalization.

In section 4, I note that traditional consequentialist rules conflict not only with conditionalization but also with probabilism. Caie (2013) has argued on these grounds that it can be rational to have probabilistically incoherent credences. I suggest, by contrast, that these examples show that rationality is not governed by consequentialist rules. I also argue that focusing on the wrong type of example has made it easy to overlook the real import of dependency relations between acts and accuracy.

In section 5, I argue that while nonconsequentialist rules can generate more intuitive results than consequentialist rules (in that they preserve plausible coherence norms), nonconsequentialist rules come at a cost: they don't have the same intuitive justification as consequentialist rules. We cannot use the idea that belief aims at the truth to motivate nonconsequentialist rules, as we could with consequentialist rules. In section 6, I consider and reject the argument that N-dominance avoidance derives its normative force from the normativity of logic. I conclude in section 7.

1 Epistemic utility theory

A common way of modeling belief states is to treat beliefs as coming in degrees. Instead of a tripartite division of belief, disbelief, and suspension of judgment, we represent belief states with credences in the interval from 0 to 1, where 1 represents certain belief and 0 represents certain disbelief. Total belief states are formally represented as functions from propositions (sets of possible worlds) to real values in the $[0,1]$ interval. Credences are typically held to be regulated by two coherence norms: probabilism and conditionalization.

Probabilism: a rational agent's credences form a probability function.[3]

Conditionalization: let E be the most informative proposition that an rational agent learns between t and t'. Then the agent's credences update such that $Cr_{t'}(\cdot) = Cr_t(\cdot \mid E)$.[4]

The most familiar justifications for these norms are Dutch book arguments, which purport to show that agents with credences that violate either probabilism or conditionalization are susceptible to foreseeable exploitation. (For example, they'll accept sets of bets that the agent is in a position to know will jointly guarantee a sure loss.) But many have thought that this sort of justification is inadequate. We have good epistemic reason to have coherent credences, but Dutch book arguments give only pragmatic reasons for coherence.[5]

Epistemic utility theory aims to provide non-pragmatic justifications for epistemic norms including probabilism and conditionalization. The basic tools of decision theory are given an epistemic flavor: the relevant sorts of "acts" are epistemic, i.e. the (presumably involuntary) adoption of credence functions,[6] and the relevant sort of value measured by the utility function is epistemic value. What do epistemic utility functions look like? The natural candidates are those utility functions that tie credences to the ideal of *truth*. A credence has greater epistemic utility at a world if it's closer to the truth in that world, i.e., if it has greater accuracy.

Joyce (1998) offers a non-pragmatic justification for probabilism. His argument appeals to a decision rule that Joyce calls "Dominance." Dominance prohibits adopting a credence function if (i) another is more accurate at every possible world and (ii) there is some credence function that is not dominated. Joyce then proves that for a range of epistemic utility functions satisfying certain constraints, any credence function that violates probabilism is strictly dominated.

Instead of dominance avoidance, Greaves and Wallace (2006) and Leitgeb and Pettigrew (2010a,b) argue for coherence norms by appeal to expected epistemic utility max-

---

[3] That is, they obey the following three axioms. Where $\mathscr{W}$ is the (finite) set of all worlds under consideration:

1. *Nonnegativity:* for all propositions $X \subseteq \mathscr{W}$, $Cr(X) \geq 0$
2. *Normalization:* $Cr(\mathscr{W}) = 1$
3. *Finite additivity:* if X and Y are disjoint, then $Cr(X \cup Y) = Cr(X) + Cr(Y)$

[4] Where $Cr(X \mid Y)$ is standardly defined as $\frac{Cr(X \cap Y)}{Cr(Y)}$.

[5] See, e.g., (Kennedy and Chihara, 1979), (Rosenkrantz, 1981), (Christensen, 1996), and (Joyce, 1998).

[6] Talk of epistemic "options" or "acts" makes it sound as though epistemic utility theory is committed to doxastic voluntarism. But these descriptions are only meant to make the analogy with practical decision theory explicit. The project can be thought of as a system of third-personal evaluations rather than first-personal guides to "action," and so there's no more need for doxastic voluntarism in epistemic utility theory than anywhere else in epistemology.

imization. Expected utility is the average utility of a set of possible mutually exclusive outcomes, weighted by some relevant measure of probability. Greaves & Wallace argue that, upon receiving new evidence, the update procedure that uniquely maximizes expected accuracy is conditionalization on the new evidence. Leitgeb & Pettigrew provide expected accuracy arguments for both probabilism and conditionalization.

What all of these arguments have in common is that they aim to derive results simply by plugging epistemological tools into the familiar decision-theoretic apparatus: utility functions that encode what's epistemically valuable (accuracy) and epistemic "acts" (adopting credences).

But in fact, the rules that epistemic utility theorists have appealed to differ in surprising ways from those used in practical decision theory. These rules use different notions of expected utility and dominance. In the epistemic utility theoretic arguments for probabilism and conditionalization cited above, the differences haven't been acknowledged or justified. As I'll show, the differences aren't innocuous. First, as Caie (2013) and Greaves (2013) show, they produce very different results; different epistemic actions are sanctioned. Second, I'll argue, they require a very different philosophical interpretation: they cannot be applied to a person's credal acts or credal states.

2 How epistemic acts affect their own accuracy

2.1 The central example

Greaves and Wallace (2006) and Leitgeb and Pettigrew (2010b) argue that expected accuracy maximization requires agents to conditionalize on any new evidence:

Expected accuracy maximization: it's irrational to have an update policy that fails to maximize expected accuracy (from the agent's perspective).

I'll argue that on the most natural interpretation of this rule, it conflicts with conditionalization.

The kind of cases where expected accuracy maximization can conflict with conditionalization are cases where credal acts are not causally or evidentially independent of events in the world.[7] There are a variety of such cases. For example: your credence in a proposition can causally affect the likelihood that the proposition is true: a belief might be self-verifying,[8] self-falsifying,[9] etc. Your credence in a proposition can also causally affect the

[7]See (Berker, 2013b), (Caie, 2013), (Greaves, 2013).

[8]William James discusses cases like this as examples of practical reasons for belief: "There are then cases where faith creates its own verification. Believe, and you shall be right, for you shall save yourself; doubt, and you shall again be right, for you shall perish" (James, 1896, 96–7).

[9]The TV show *Arrested Development* provides one such example: an out-of-work actor auditioning for the role of a frightened inmate is more likely to get the role if he believes he won't (because he'll exhibit genuine fear) and less likely to get the role if he believes he will (because he'll fail to exhibit genuine fear). (Season 1, episode 3: "Marta Complex.")

likelihood that other propositions are true.[10] Your lower-order credences determine the accuracy of your higher-order credences. Similarly for other mental states: your credences about what you will do are at least not evidentially independent of what you will do.[11] Your credences are not evidentially independent, and arguably not causally independent, of others' credences, including their credences about your credences, e.g., in communication. Similarly, your credences are not evidentially independent of other facts about the world, e.g. others' actions when you believe you'll command them. Your credences in certain propositions can logically determine whether those propositions are true.[12] Your credences may determine the likelihood of certain kinds of social facts: for example, if you're The Trendsetter and you believe something is cool, that can make it the case that it is cool (constitutively, rather than causally). And so on.

In short: which world you're in is partly dependent on your epistemic acts. And so your epistemic acts can influence their own degree of accuracy. We need to understand the role of this kind of dependency in epistemic utility theory.

First, let's examine the effects of this kind of dependency on the standard machinery of epistemic utility theory. I will be focusing on examples of the first and second kind: that is, beliefs, or pairs of beliefs, that are self-verifying or self-falsifying.[13] In the following example, an agent's credence in a proposition causes that proposition to be more or less likely to be true. I'll return to this example throughout this paper.

> **Example #1**: Handstand
>
> Suppose your (perfectly reliable) yoga teacher has informed you that the only thing that could inhibit your ability to do a handstand is self-doubt, which can make you unstable or even hamper your ability to kick up into the upside-down position. The more confident you are that you will manage to do a handstand, the more likely it is that you will, and vice versa.
>
> More precisely: let H be the proposition that you'll successfully do a handstand at t_3. Your yoga teacher has informed you that for all n in $[0,1]$, $Cr(H) = n$ at

[10] See (Berker, 2013a) for examples and discussion.

[11] Typically you believe you'll do what you intend to do; this belief is, on some views, a necessary condition on the intention. See Audi (1973); Harman (1997); Davis (1984); Velleman (1989); Ross (2009); Sinhababu (2013).

[12] For example, in the case of certain self-referential beliefs: the proposition *I believe this proposition is true* is true just in case I believe that proposition; by believing it I make it true. For a discussion of epistemic rationality and propositions of this sort, see (Caie, 2012) and (Caie, 2013).

[13] Examples of the first and second kind are central to Berker's (2013) argument against various forms of epistemic consequentialism, such as reliabilism. Berker presents the following scenario: suppose an agent's believing that she will recover from a particular illness makes it the case that she will recover from the illness, but she has evidence that 80% of people with this illness do not recover. Is she epistemically required to believe that she will recover from the illness, given that doing so is conducive to having true belief? Intuitively not. The focus of Berker's argument is various forms of *objective* epistemic consequentialism. Berker doesn't direct his argument explicitly toward epistemic utility theory, which at first glance seems to be a form of *subjective* epistemic consequentialism. As we'll see, constructing examples of causal dependency between belief and truth for epistemic utility opens up foundational questions about what sort of project epistemic utility theory is really engaged in, and whether it is genuinely a form of subjective epistemic consequentialism.

> t_1 will make it the case that $Ch(H) = n$ at the next moment, t_2, and that this will remain the chance of H up to t_3, when you either do or don't do a successful handstand. We'll call the information she's given you: "$Cr(H) = Ch(H)$," where "Cr" and "Ch" nonrigidly pick out, respectively, whichever credence function you adopt at t_1 at a world and the chances at t_2 at that world.
>
> Suppose that in the actual world, your prior Cr_0 is such that before learning $Cr(H) = Ch(H)$, your conditional credence in H given that information is .5:
>
> $$Cr_0(H \mid Cr(H) = Ch(H)) = .5$$

Let's assume for the moment that this is a rational credence to have. After all, the result that conditionalization always uniquely maximizes expected accuracy is supposed to hold for *any* probabilistic priors. So if there is a divergence between conditionalization and expected accuracy maximization, then it doesn't matter whether these priors are rational, so long as they're probabilistic. (Furthermore, as we'll see, there are good reasons for allowing these priors to be rational.)

What credence is epistemically rational for you to adopt at t_1 upon learning $Cr(H) = Ch(H)$?[14] Conditionalization, of course, says that your updated credence should be .5. The question is, what do expected accuracy maximization and dominance avoidance say? We'll turn to expected accuracy maximization first and return to dominance avoidance in section 3.[15]

2.2 Two kinds of decision theory

In order to know what expected accuracy maximization recommends, we need to specify what the rule says. Expected accuracy maximization is a special case of a decision-theoretic rule, expected utility maximization, paired with an accuracy-based utility function. There are two competing forms that expected utility maximization most commonly takes in decision theory: causal expected utility maximization and evidential expected utility maximization. As we will see, both of these rules differ in interesting ways from the nonconsequentialist decision rules used in epistemic utility theory, and as a result they make different predictions from the decision rules epistemic utility theorists use. But because epistemic utility theorists have mostly ignored this difference, it's helpful to see what would happen if

[14] This example has a confounding factor: that it's desirable to be able to do a handstand. This makes the idea that one should have credence 1 in H appealing from a pragmatic perspective. Indeed, James (1896) discusses examples of this sort of belief efficacy as a way of arguing that there can be practical reasons for belief. But my discussion concerns with epistemic rationality, not practical or all-things-considered rationality.

[15] Berker (2013b) and Greaves (2013) both discuss self-verifying beliefs; Greaves discusses a case similar to Handstand and notes that both causal and evidential decision theory make the predictions I'll spell out in the next two sections. Greaves claims that both causal and evidential decision theory make the correct predictions about self-verifying cases, capturing the "intuition" that "only extremal credences are rationally permitted, since these and no others lead with certainty to perfect match between ones credences and the truth" (932). Greaves accepts as an intuition a claim that this paper centrally denies. In section 2.5, I'll provide three arguments that, in fact, evidential and causal decision theory make the wrong predictions about the case.

we simply took traditional practical decision theories and fed into them an epistemic utility function.

The **causal expected utility** of an act is a measure of the value you can expect to result from taking that act. Here's a natural way of calculating causal expected utility (from Lewis 1981): we partition the space of possibilities into a set of so-called "dependency hypotheses." A dependency hypothesis is a maximally specific proposition about how facts about the world causally depend on the agent's present acts. The agent might be uncertain about what effects her acts can have, and so there may be many epistemically possible dependency hypotheses. The causal expected utility of an act is the weighted average of the value of each possible causal outcome of the act, weighted by the probability of the dependency hypothesis where the act causes that outcome. Formally, the causal expected utility is calculated as follows: where each k_i is a possible dependency hypothesis, $CEU(a) = \sum_i Cr(k_i)U(k_i \wedge a)$.

The **evidential expected utility** of an act differs from its causal expected utility in roughly this way: it doesn't distinguish between cases where your acts cause certain outcomes and cases where your acts merely correlate with those outcomes (perhaps because, e.g., they have a common cause, or perhaps just by chance). The evidential expected utility of an outcome is calculated as follows: where each s_i is a possible state of the world (such that the set of states forms a partition), $EEU(a) = \sum_i Cr(s_i \mid a)U(s_i \wedge a)$.

There are other forms of expected utility theory, but these two are the most familiar forms and I'll focus on them in the remaining discussion.

2.3 Causal expected accuracy

Returning to our example: for simplicity's sake, let's suppose once you've learned that your credence in H determines the chance of H such that $Cr(H) = Ch(H)$, there are only three options for which credence function to adopt[16]:

$$Cr_a(H) = 1 \qquad\qquad Cr_b(H) = .5 \qquad\qquad Cr_c(H) = 0$$

We'll assume that the only credence functions you can adopt are probability functions that give credence 1 to $Cr(H) = Ch(H)$. So once we stipulate that the only propositions these credence functions are defined over are H, propositions concerning which credence function you have, propositions concerning which chance function obtains, and the Boolean closure of all of them, then this is all we need to fully specify each credence function.

In our example, then, we only have to consider two dependency hypotheses, which differ only in whether H would be true if you were to adopt Cr_b:[17]

[16] I'll leave it as an exercise to the skeptical reader to fill in the remaining options $Cr(H) \in (0, .5) \cup (.5, 1)$; these other options won't make any difference to my point.

[17] Notation: "$\overline{H}$" is the negation of H. "$\Box\!\!\rightarrow$" is the counterfactual connective: *if...would...*. I've represented the dependency hypotheses as including determinate facts about H. Lewis (1981), by contrast, does not distinguish dependency hypotheses that differ only with respect to chances, and so would treat this as a case where there's only one dependency hypothesis with positive probability. For assigning utilities in terms of gradational accuracy (as opposed to gradational proximity to chances, a possibility I discuss later), we need the non-Lewisian tweak: we can't say how accurate a credence in H is at a state unless we specify whether H

$$k_1: \text{ you adopt } Cr_a \;\Box\!\!\rightarrow H \qquad\qquad k_2: \text{ you adopt } Cr_a \;\Box\!\!\rightarrow H$$
$$\text{you adopt } Cr_b \;\Box\!\!\rightarrow H \qquad\qquad \text{you adopt } Cr_b \;\Box\!\!\rightarrow \overline{H}$$
$$\text{you adopt } Cr_c \;\Box\!\!\rightarrow \overline{H} \qquad\qquad \text{you adopt } Cr_c \;\Box\!\!\rightarrow \overline{H}$$

The reason we only have these two is that we're including the information that $Cr(H) = Ch(H)$. Other dependency hypotheses where, e.g., $Cr_a \;\Box\!\!\rightarrow \overline{H}$ are all given probability zero, and so we don't have to consider them.

So there are two states, k_1 and k_2, and three options. This gives us a partition over epistemic possibilities; we can treat the cells of the partition as worlds. We then calculate the accuracy of each epistemic act at all worlds where that act is taken.

We're interested in the local accuracy of your credence in a particular proposition. A **local accuracy** measure characterizes the accuracy of an agent's credence in a particular proposition. A **global accuracy** measure characterizes the accuracy of a total credence function. We can make a minimal assumption about accuracy measures: that credence 1 in a truth and credence 0 in a falsehood have maximal accuracy, and any intermediate credence has some positive distance from the truth, hence does not have maximal accuracy.

	k_1	k_2
Cr_a	maximal	maximal
Cr_b	not maximal	not maximal
Cr_c	maximal	maximal

There's no need to calculate the causal expected accuracy because Cr_b is strictly dominated.[18] So we know that Cr_b does not maximize causal expected accuracy.

Causal expected accuracy (CEA) maximization therefore conflicts with conditionalization. Conditionalization requires you to adopt Cr_b, whereas maximizing CEA requires you to adopt either Cr_a or Cr_c.

2.4 Evidential expected accuracy

What about evidential expected accuracy (EEA) maximization? First, note that only four possibilities (w_1, w_2, w_3, w_4) have positive probability (ex hypothesi).

is true at that state.

[18]By "dominated," here I mean the term in the sense used in practical decision theory: that there's some alternative act such that for every state (that is, dependency hypothesis), the outcome of Cr_b is worse than every outcome of performing that act.

	Ch_a, H	Ch_b, H	Ch_c, H	$Ch_a, \overline{H}$	$Ch_b, \overline{H}$	$Ch_c, \overline{H}$
Cr_a	w_1	$\diagup$	$\diagdown$	$\diagdown$	$\diagup$	$\diagup$
Cr_b	$\diagup$	w_2	$\diagdown$	$\diagdown$	w_3	$\diagup$
Cr_c	$\diagup$	$\diagup$	$\diagdown$	$\diagdown$	$\diagup$	w_4

$\diagdown$ = ruled out by chance info; $\diagup$ = ruled out by $Ch(H) = Cr(H)$

Notation: "Cr", in sans serif, denotes the proposition that you adopt the credence function Cr; "w", in sans serif, denotes the maximally specific proposition that is only true at w.[19]

Again, we assume that our epistemic utility function is such that only credence 1 in truths and credence 0 in falsehoods have maximal utility.

$$EEA(Cr_a(H)) = Cr_0(\mathsf{w_1} \mid \mathsf{Cr_a})U(Cr_a, w_1)$$
$$EEA(Cr_b(H)) = Cr_0(\mathsf{w_2} \mid \mathsf{Cr_b})U(Cr_b, w_2) + Cr_0(\mathsf{w_3} \mid \mathsf{Cr_b})U(Cr_b, w_3)$$
$$EEA(Cr_c(H)) = Cr_0(\mathsf{w_4} \mid \mathsf{Cr_a})U(Cr_c, w_4)$$

The evidential expected accuracy of $Cr_a(H)$ and $Cr_c(H)$ will both be maximal; the evidential expected accuracy of $Cr_b(H)$ will be a weighted average of two nonmaximal utilities, therefore nonmaximal. So whatever your prior credence function (Cr_0) is, as long as it's probabilistically coherent and updated on $Cr(H) = Ch(H)$, it will give Cr_b lower evidential expected accuracy than Cr_a or Cr_c. And the same holds for any alternative to Cr_a or Cr_c that assigns non-extremal value to H, and for any epistemic utility function that assigns higher utility to credence that match truth values than credences that don't.

So no matter what your prior is, EEA maximization instructs you to adopt extremal credences. And this means that, like CEA maximization, it conflicts with conditionalization in all cases where $Cr_0(H \mid Cr(H) = Ch(H))$ is not extremal. It's no surprise that causal and evidential decision theories agree in this case: the dependency relation between H and your credence in H is causal as well as evidential.

2.5 Discussion

Greaves and Wallace (2006) and Leitgeb and Pettigrew (2010b) argued that planning to

[19] Note: $U(Cr, w)$ doesn't use the sans serif notation because U's arguments are a credence function and a world, not a conjunction of propositions. The reason for this is that any credence function can be paired with any world as input to U, even when $\mathsf{Cr} \cap \mathsf{w} = \varnothing$ (i.e. whenever w happens to be a world where you adopt a different credence function, or don't exist at all).

update by conditionalization uniquely maximizes expected accuracy, from the perspective of any probabilistic priors. I've shown that if expected accuracy maximization is cashed out evidentially or causally, this is not true.

Now, one might concede this point and be willing to give up on the general result—but still want to embrace both conditionalization and causal or evidential expected accuracy. That is, one might accept that there can be conflicts between conditionalization and causal and evidential expected accuracy maximization for some probabilistic priors, but still accept both norms as constraints on rational agents. And there is a fall-back position that allows one to do so in the example I gave: by ruling nonextremal prior credences in H conditional on $Cr(H) = Ch(H)$ irrational. Thus, any rational agent updating by conditionalization on $Cr(H) = Ch(H)$ will end up with a credence in H that maximizes causal and expected expected accuracy. We can't get the right results for *all* probabilistic credences, one might say; but by being less permissive priors, we can at least get them for the rational ones.

But as I'll argue, this strategy has costs. The handstand example, and other cases of act-dependence in epistemic utility theory, reveal that the goal of having accurate credences is different from, and can even conflict with, the goal of having the credences that are supported by the evidence. I'll discuss this conclusion more in section 4.1.[20]

3 Nonconsequentialist rules

3.1 Consequentialist and nonconsequentialist expected accuracy

Now, you might think I have provided a counterexample to the arguments in Greaves and Wallace (2006) and Leitgeb and Pettigrew (2010b) that expected accuracy maximization entails planning to update by conditionalization. But I haven't. What I've shown is that if we conceive of expected accuracy in the common decision-theoretic way—as something like causal or evidential expected accuracy—then we can't get this result.

Evidential expected accuracy: $EEA_{Cr}(Cr') = \sum_{w \in W} Cr(\mathsf{w} \mid \mathsf{Cr'})U(Cr', w)$

Causal expected accuracy: $CEA_{Cr}(Cr') = \sum_{k \in K} Cr(k)U(Cr', k)$

Call CEA maximization and EEA maximization **consequentialist** rules: they are concerned with the outcomes of credal acts. Traditional epistemic utility theorists (Joyce, 1998, 2009, Greaves and Wallace, 2006, Leitgeb and Pettigrew, 2010a,b), use a notion of expected accuracy different from either causal or evidential expected accuracy. We'll call their inaccuracy measure "nonconsequentialist expected accuracy" or "NEA." Where E is the set of epistemically possible worlds,

Nonconsequentialist expected accuracy: $NEA_{Cr}(Cr') = \sum_{w \in E} Cr(\mathsf{w})U(Cr', w)$

NEA maximization does recommend the same epistemic acts as conditionalization.

[20]See (Berker, 2013a) and (Berker, 2013b) for a more thorough discussion of the difference between evidential support and truth-conducivity.

How? NEA doesn't take into account, in its calculation of the expected accuracy of an epistemic act, the fact that the epistemic act is taken. That is, it doesn't take into account any dependency relations, causal or evidential, between adopting particular credences and what the world is like. By contrast, evidential expected accuracy weights the utility of an epistemic act at different worlds by the probability of that world *conditional on the performance of that epistemic act.* Causal expected accuracy uses partitions of worlds that are based on the *effects* epistemic acts might have. In both cases, the weighting on the utilities for the expectation depends on which act is performed in the worlds considered.

3.2 Consequentialist and nonconsequentialist accuracy dominance

There is a similar distinction to be drawn for accuracy dominance avoidance: it comes in consequentialist and nonconsequentialist varieties. The consequentialist version is better at characterizing the aim of accuracy, but it conflicts with coherence norms like conditionalization and probabilism.

The relation that Joyce calls "dominance" in (Joyce 1998, 2009) is what I'll call "nonconsequentialist dominance" or "N-dominance." Here's how Joyce defined the relation of N-dominance:

> Cr strictly **N-dominates** Cr' relative to utility function U iff $U(Cr,w) > U(Cr',w)$ for all $w \in W$.

By contrast, here is the standard, consequentialist definition of dominance in practical decision theory (what I'll call "C-dominance"):

> Cr strictly C-**dominates** Cr' relative to U iff $U(\mathsf{Cr} \cap s) > U(\mathsf{Cr'} \cap s)$ for all $s \in S$.[21]

The difference: N-dominance, unlike C-dominance, takes into account the value of acts at *all* worlds—including worlds where the acts under assessment are not performed.

If we use C-dominance, then there will be some probabilistically coherent credence functions that are dominated. Furthermore, it will no longer be obvious that all credences that are accuracy-C-dominated are irrational.

Consider, again, the self-fulfilling example. Since all of the credence functions we considered were probabilistically coherent, none were N-dominated.[22] But Cr_b—along with any other non-extremal credence in H—will be C-dominated. And so they'll be ruled epistemically impermissible. So if we move to consequentialist dominance avoidance, we will

[21] Note that how exactly to partition states for the purposes of epistemic utility theory is not obvious: we must build in enough information that it's possible to assess credence functions for accuracy, but not so much information that the acts and states are no longer orthogonal. One possibility is to use dependency hypotheses as states; presumably there are other ways of determining adequately fine-grained partitions. Thanks to Catrin Campbell-Moore for pointing out this challenge.

[22] That is, on certain assumptions about the choice of epistemic utility function. Joyce shows this in his (2009).

find conflicts with conditionalization. Credence functions that seem perfectly reasonable will be ruled rationally impermissible.

The difference between nonconsequentialist rules and traditional, consequentialist rules has gone largely unnoticed. Those who have noticed this difference (Greaves 2013 and Caie 2013) have treated the nonconsequentialist rules as convenient simplifications, not competitors to consequentialist rules. They have argued that we should "fix" epistemic utility theory by replacing nonconsequentialist rules with consequentialist rules.

I think this is a mistake. Rational belief conforms itself to the evidence: it aims to match the world, not to mold the world so that it matches belief. And so, as I'll argue in much of the rest of the paper, NEA's failure to consider the consequences of belief is a feature, not a bug.[23]

4 Coherence vs. accuracy

4.1 Evidence vs. accuracy

Here are two claims that are pretty plausible:

Evidence You should adopt the belief state that your evidence supports.[24]

[23] A switch from NEA to CEA or EEA has wide-ranging effects on epistemic utility theory: not only can we not derive the same conclusions, but we can't even rely on the same premises. A standard starting point in epistemic utility theory is to characterize plausible epistemic utility functions. A constraint on epistemic utility functions, Joyce and Leitgeb and Pettigrew defend, is that epistemic utility functions must be proper:

Propriety: Any probabilistic credence assigns itself maximal expected accuracy.

As Joyce notes, a variety of different epistemic utility functions satisfy this desideratum, including quadratic scoring rules (like the Brier score) and logarithmic scoring rules. This is so, though, only with NEA. What happens if we use CEA or EEA?

In the handstand case, it's straightforward that no credence functions that assigns a credence in the open $(0,1)$ interval to H maximizes CEA by its own lights, since all such credence functions are dominated. The argument involving for the same conclusion about EEA is more complicated. Call a credence function "transparent" iff it conforms to to the following rule:

Transparency: $Cr(H) = n$ only if $[Cr(\mathsf{Cr(H) = n}) = 1$ and for all $m \neq n$, $Cr(\mathsf{Cr(H) = m}) = 0]$.

With CEA and EEA, only credence functions that are transparent can maximize causal or evidential expected accuracy. So, we'll restrict ourselves to transparent credence functions. But then the EEA of all credence functions other than one's own will be undefined. After all, $EEA_{Cr}(Cr')$ is calculated using $Cr(\mathsf{w} \mid \mathsf{Cr'})$. But since Cr must be transparent, it assigns $\mathsf{Cr'}$ probability zero, and so the conditional probability will be undefined. And the EEA_{Cr} of Cr won't be higher than any other credence function: all alternatives will have undefined EEA_{Cr}. So relative to CEA, no accuracy measure will satisfy propriety. And relative to EEA, propriety is trivial, since the EEA of any credence function other than the assessor's is undefined.

[24] Many consider this norm a platitude (Williamson, 2000; Kelly, 2014) or as definitional (Kim, 1988); but it is not uncontroversial. For a canonical defense, see (Feldman and Conee, 1985). For dissenting views, see e.g. (James, 1896; Goldman, 1979; Wedgwood, ms).

Accuracy What you should believe is determined by the aim of accuracy: that is, the aim of having a belief state that's as close to the truth as possible.

These claims seem to go tidily together. It can even seem like Evidence and Accuracy are two ways of expressing the same thing. After all, the evidence supports believing that H only if the evidence suggests that H is true. That's what makes it *evidential* support. Some considerations might favor believing what's false, maybe because you derive some benefit from having a false belief. But evidential considerations in favor of a belief are considerations that point you in the direction of the truth, or at least what your evidence suggests is true. If you had perfect, complete evidence, then it would support perfectly accurate beliefs.

Now, in the handstand case, conditionalization permits extremal and non-extremal credences in H (depending on priors). It's clear, however, that Accuracy permits only extremal credences. After all, a policy that required the adoption of extremal credences would guarantee perfectly accurate credences; any other policy would guarantee credences that were not perfectly accurate. Does Evidence yield the same verdict as Accuracy?

No: the handstand case shows that the norms of Evidence and Accuracy conflict. Here are three arguments that the Evidence norm permits non-extremal credences in H.[25]

First, it's implausible that an ideally rational agent could never have good evidence for thinking: "I'm just the kind of person who'll be uncertain about H in circumstances like this; and so there's no telling whether H will be true." Perhaps a rational agent might believe this on the basis of self-observation and induction. Perhaps she might believe it on the basis of expert testimony. Neuroscientists might inform the rational agent that brain scans make it clear that she has just this disposition. It would be surprising if rationality required us to ignore all of these canonically reliable sources of evidence.

Second, learning $Cr(H) = Ch(H)$ doesn't (necessarily) give any information in favor of or against H. Indeed, it's perfectly symmetrical with respect to whether H is true. When the evidence is perfectly symmetrical with respect to whether a hypothesis is true or false—when it gives no reason to suppose the hypothesis is true rather than its negation—many find it intuitive that the rational response is to adopt a credence that is correspondingly symmetric with respect to its level of commitment to the hypothesis vs. its negation.[26] This sort of response should be at the very least rationally permissible. So why would it be irrational to have credence .5 in H, which is also symmetrical about whether H is true?

Third, the Accuracy norm, and its consequentialist decision rules, permit only credence 1 or 0 in H. But it would be strange to think that the evidence supported our hypothesis H *both* to degree 1 and to degree 0—but nothing in between! One might find it peculiar to think that an agent's evidence could support both believing a hypothesis with certainty and

[25] The idea that epistemic utility theory might be in tension with other forms of evidentialist norms, in particular the Principal Principle, has received some recent attention. Easwaran and Fitelson (2012) argue that (one form of) the dominance avoidance principle interacts with the principal principle in a way that generates an undesirable order-dependence. Joyce (ms) and Pettigrew (2013) argue that the worry is avoidable if we either switch to a weaker form of dominance avoidance or to objective expected accuracy.

[26] Various formulations of the Principle of Indifference entail this.

rejecting it with certainty. But even if we accept that the evidence can behave in that way, there's still something odd about the idea that a single body of evidence could point toward full belief in the hypothesis and also point toward full rejection of the hypothesis, without supporting any intermediate degree of belief.[27]

4.2 Probabilism, dominance, and expected accuracy

I've claimed that epistemic utility theory faces a dilemma: we must either adopt a consequentialist decision theory that reflects how actions can affect the world (thereby giving up conditionalization and probabilism), or else give a nonconsequentialist justification for nonconsequentialist decision theory (thereby giving up the idea that rational belief aims for accuracy). To say that maximizing NEA or avoiding N-domination ensures that our beliefs are closer to the truth can't be right. The handstand case is a counterexample: a rational agent can be certain that a particular pair of credences are closest to the truth, even though those credences are not recommended by N-dominance avoidance and are ruled out by NEA maximization.

Caie (2013) and Greaves (2013) accept the first horn of the dilemma. If rationality requires aiming for true belief, then we should abandon nonconsequentialist rules. But I tollens where they ponens. I think examples like the handstand case show that rationality doesn't require us to aim for true belief.

Before arguing for the second horn, it's useful to summarize the opposing position. For simplicity I'll focus on Caie's version, because Greaves' position is somewhat more complex. Caie (2013) uses an accuracy C-dominance argument to defend rational violations of probabilism. His preferred examples involve self-referential beliefs, but as Caie notes, can be generalized to causal examples.

> **Example #2**: Frustrating Handstand
>
> Your perfectly reliable yoga teacher informs you at t_0 that at a future time t_1, your credence in H will causally determine whether H is true: H will be false if and only if you are more confident than not in H at t_1:

[27] A natural thought: maybe the problem is that we're using *accuracy*—distance from truth—to generate our epistemic utility function. What if instead of measuring credences' utility by distance from *truth*, we measure them by distance from *objective chance*? Then whatever your credence in H is, it'll have zero distance from the chance; all options are equally good. So we end up with a maximally permissive recommendation for H: any credence in H you adopt will match the objective chance, so it's rationally permissible to adopt any credence. And maybe that's an attractive outcome in this case.

Even if you think this is the right result for our first example, there are problems for this account. First, it still doesn't vindicate conditionalization, since it permits you to update to credences that aren't permitted by conditionalization. Second, the solution doesn't generalize. Consider another example: suppose you learn that whatever credence you adopt in H will make it the case that the objective chance of H is .5 lower or higher—*unless* you adopt credence .79 in H, in which case $Ch(H) = .79$. In these circumstances, adopting $Cr(H) = .79$ will maximize expected proximity to objective chance. But this is no reason to think that the evidence supports H to degree .79.

In this paper I'll be mainly focusing on accuracy measures. But it's clear that other types of epistemic utility function could be used to generate a variety of interesting results.

$$H \text{ is true iff } Cr(H) \leq .5.$$ [28]

> At t_1, what credence should you have in H, and in not-H, given that $H \leftrightarrow Cr(H) \leq .5$?

The credence in H that is not C-dominated and that maximizes CEA and EEA is .5. Adopting that credence makes H true; and so in this situation, the *pair* of credences in H and not-H that both C-dominates and maximizes CEA and EEA is $Cr(H) = .5$ and $Cr(\overline{H}) = 0$. This pair of credences is probabilistically incoherent.

So consequentialist decision rules (accuracy C-dominance avoidance, CEA maximization, EEA maximization) will sometimes recommend violating probabilism. Caie (2013) argues that since there are circumstances where we can get our credences closest to the truth by violating probabilism, we sometimes rationally ought to violate probabilism.

In self-frustrating cases like example #2, there are no obviously attractive options. One can either have nonprobabilistic credences—which many find intuitively irrational—or one can have probabilistic but unstable credences. (In example #2, any probabilistic credences will be unstable because once you know what credence you've adopted, the appropriate response to that information is to infer from it whether H will be true.) Since both options are relatively unintuitive, the example doesn't push in one direction or another.

On the other hand, if we look to self-fulfilling cases like my example #1, the balance is broken. There's nothing intuitively epistemically irrational about being uncertain whether H in that situation. Indeed, it's odd to think that absolute certainty in H or not-H is rationally *required* in response to evidence like $Cr(H) = Ch(H)$, which is entirely symmetrical with respect to whether H is true.

There are other benefits to looking at self-fulfilling cases rather than self-frustrating cases: they cannot be explained away as easily.

4.3 Can distinguishing different loci of evaluation explain away the worry?

Joyce (2013) argues that cases like example #2 can be explained away while retaining a consequentialist interpretation of epistemic utility theory. Rather than revealing counterexamples to probabilism (as Caie maintains) or to the consequentialist interpretation of epistemic utility theory (as I am arguing), Joyce thinks such examples can be taken care of by appeal to different loci of evaluation. We should evaluate epistemic "choices" or acts differently from how we evaluate resulting epistemic states.

Joyce offers an analogy: according to the causal decision theorist, if you are offered a pill that makes you a one-boxer, you should take it. But if, as a result, you one-box in a Newcomb situation, the causal decision theorist will evaluate you as having acted irrationally. Similarly, it might be that we should advise someone in example #2's unfortunate

[28] "Cr" is a definite description of your credence function at t_1.

situation to take the following credal act: adopt $Cr(H) = .5$ and $Cr(\overline{H}) = 0$. But we can nevertheless evaluate the resulting state negatively, for it is accuracy N-dominated.[29]

However plausible we might find this explanation with respect to self-frustrating cases like example #2, again, it's helpful to consider self-fulfilling cases like example #1. Here the analogy is less compelling. In the one-boxer-pill case, there are two distinct loci of evaluation: the act of taking the pill (which maximizes causal expected utility) and the later act of one-boxing (which does not maximize causal expected utility). Similarly for example #2: it might be at least prima facie plausible that, in example #2, we evaluate the act (i.e. the adoption of certain nonprobabilistic credences) positively, and yet evaluate the resulting epistemic state negatively. Again, there are two distinct loci of evaluation.

But example #1, the self-fulfilling case, does not have that structure. There, the attitude that gets you closest to the truth—adopting either $Cr(H) = 1$ or $Cr(H) = 0$—will be probabilistic. Indeed, considered on its own merits, it's entirely unproblematic. The only negative thing to be said about the attitude that results from adopting credence 1 or 0 in H is that it is the result of a bad update policy (because it involves a violation of conditionalization). But then it doesn't make sense to also recommend that very same update policy (because it maximizes accuracy and expected accuracy). That would involve conflicting assessments of one and the same locus of evaluation.

In short: we do need to make a choice between coherence norms, like conditionalization and probabilism, and consequentialist rules, like accuracy C-dominance avoidance and CEA/EEA maximization. If we give up consequentialist rules, as I recommend, then we can no longer think of epistemic utility theory as codifying the aim of accuracy.

5 The aim of accuracy

5.1 Evaluating credal acts at worlds where they aren't performed

The difference between consequentialist rules and nonconsequentialist rules is this: consequentialist rules evaluate each epistemic act based on the value of its possible outcomes (construed non-causally), that is, how close some body of beliefs will be to the truth if it is adopted. Nonconsequentialist rules evaluate epistemic acts based on something more: they look at the "value" of each epistemic act at *all* possible worlds, even those where the act isn't performed.

This is puzzling. What is the value of adopting a credence at an outcome where no one adopts the credence function? It doesn't make sense to talk of an act's *value* at a world where it isn't performed. Certainly, it makes sense to talk of the closeness of a credence function, understood as a mathematical object, to the truth values at a world where you don't adopt that credence function. But we cannot interpret that as the *epistemic utility of*

[29] Neither Joyce nor Caie discusses the distinction between C-dominance and N-dominance, but we shouldn't interpret them as endorsing incompatible conclusions with respect to whether that pair of credences is dominated. There is a non-probabilistic credence function Cr_n which C-dominates all alternatives; but it is N-dominated. That is, there is a probabilistic credence function Cr_p which N-dominates Cr_n—but only if the relevant agent doesn't adopt Cr_p.

adopting that credence function at that world, or the *accuracy of your doxastic state* at that world.[30] And so we cannot interpret nonconsequentialist rules as means to an epistemic end.

Consider Joyce's (1998, 2009) "dominance" argument for probabilism. (The argument is, of course, actually an N-dominance argument for probabilism, though Joyce's paper does not distinguish the two decision rules.) A premise of his argument is that it is irrational to adopt credences that are accuracy N-dominated. But the rationale for that premise equally motivates the claim that it's irrational to adopt credences that are accuracy C-dominated. Here is Joyce's (2009) justification for the claim that it's irrational to adopt N-dominated credences:

> [W]hen we endorse a rule as the correct measure of epistemic disutility we commit ourselves to thinking that there is something defective, from a purely epistemic perspective, about credences that score poorly according to that rule. Moreover, if these poor scores arise as a matter of necessity, then the defect is one of epistemic irrationality. (267)

Joyce claims that N-dominance avoidance is not a substantive thesis about epistemic rationality, but rather a constraint on what can be considered an epistemic utility function. But the claim that "if poor scores arise as a matter of necessity, then the defect is one of epistemic rationality" justifies C-dominance avoidance at least as well as N-dominance avoidance. After all, adopting C-dominated credences means adopting credences that are farther from the truth than some alternative, no matter what the rest of the world is like. And so requiring rational credences not to be N-dominated, rather than requiring them not to be C-dominated, is at the very least substantive.

So we need some justification for the claim that choosing options that are N-dominated would be irrational, given that choosing options that are C-dominated is not. And similarly, we need some justification for the claim that choosing options that are not NEA-maximal would be irrational. But this justification cannot be presented in terms of the aim of having accurate credences. If your aim is accuracy, why does it matter whether your credences are only farther from the truth than some alternative credences at worlds *where you don't have those alternative credences*?

So we're stuck with some pieces of math—the relation of N-dominance, the property of being NEA-maximal—without a clear philosophical interpretation or justification.

5.2 Credal acts and nonconsequentialist rules

I'm going to show that nonconsequentialist rules are not rules for evaluating anyone's credal acts or credal states. They only make sense interpreted as rules for evaluating credence *functions*—interpreted as mathematical objects, not mental states.

[30] Note that in causal decision theory (but not evidential decision theory), it's possible to calculate the *expected* value of an act, given that the relevant agent won't take it. But that doesn't shed any light on the notion of the value simpliciter of an act at a world where the agent doesn't take it.

First, consider the following interesting fact about nonconsequentialist expected accuracy. It is *definitional* of expected value that if the value of a random variable is known with certainty, then from the perspective of the knower, its expected value should be equal to its actual value. But this cannot be true of nonconsequentialist expected accuracy.

Let's return once more to the self-fulfilling handstand example. You know in advance that if you adopt $Cr(H) = 1$, the accuracy of your credence will be maximal. That is, you know that the credence will be perfectly accurate; its distance from the truth will 0. If you know that credence 1 in H will *in fact* have maximal accuracy, then one might expect that the *expected* accuracy you assign credence 1 in H should also be maximal. But the NEA of credence 1 in H, from your point of view, need not be maximal: given that your prior credence in H is .5, you might reasonably assign credence .5 in $\neg H$. At worlds where $\neg H$ is true, the posterior credence 1 in H is maximally *in*accurate. And so your NEA for credence 1 in H is an average of maximal and minimal utilities; hence not itself minimal. So NEA cannot be the expected accuracy of a person's credences.

Similar observations can be made about N-dominance. To visualize the contrast between C-dominance and N-dominance, consider the following familiar representation of a decision problem:

	s_1	s_2
Cr_1	w_1	w_2
Cr_2	w_3	w_4

In ordinary C-dominance arguments, we just care about whether the utility of w_1 is greater than that of w_3 and whether the utility of w_2 is greater than that of w_4. If so, Cr_1 C-dominates Cr_2. On the other hand, in N-dominance arguments, we assess whether the value of Cr_1 is greater than the value of Cr_2 at every world—that is, at all four outcomes. But this includes, e.g., w_3, a world where you adopt adopt Cr_2.

	w_1	w_2	w_3	w_4
Cr_1	$\langle Cr_1, w_1 \rangle$	$\langle Cr_1, w_2 \rangle$	$\langle Cr_1, w_3 \rangle$	$\langle Cr_1, w_4 \rangle$
Cr_2	$\langle Cr_2, w_1 \rangle$	$\langle Cr_2, w_2 \rangle$	$\langle Cr_2, w_3 \rangle$	$\langle Cr_2, w_4 \rangle$

Each option is assessed at all worlds—including worlds where it is not performed. It doesn't make sense to think of a person's credal acts or credal states in that way. We need to assess an entity that exists at each world; but there are no contingent acts that fit that description.

5.3 The aim of accuracy can't be used to motivate nonconsequentialist rules

Where the distinction between consequentialist and nonconsequentialist rules has been discussed, nonconsequentialist rules are, as a rule, treated as a naive oversimplification, to be abandoned for the more robust consequentialist rules whenever cases with more complex dependency relations arise. By and large, nonconsequentialist rules have not been treated as competing with the consequentialist rules—as alternative rules to be considered on their own merits.[31]

And so the intuitive justifications that have been given for the nonconsequentialist decision rules used in epistemic utility theory have pointed in the wrong direction. They make sense only as justifications for consequentialist decision rules: rules that require adopting the credences that best approximate the truth *in the worlds where you have those credences.* In other words, consequentialist rules are rules that require taking credal actions that can be predicted to lead to good epistemic outcomes. But nonconsequentialist rules don't always do that. In examples #1 and #2, nonconsequentialist rules permit you to adopt credences that you can be certain are farther from the truth than some possible alternatives.

Let me list some examples of considerations intended to support nonconsequentialist rules that actually support consequentialist rules, and in some cases are not even consistent with nonconsequentialist rules. Here is Joyce:

> My position is that a rational partial believer must aim. . . to <u>hold partial beliefs that are gradationally accurate</u> by adjusting the strengths of her opinions in a way that best <u>maximizes her degree of confidence in truths</u> while <u>minimizing her degree of confidence in falsehoods</u>. (Joyce, 1998, 578)

Greaves and Wallace's justification for maximizing accuracy is straightforwardly incompatible with maximizing nonconsequentialist expected accuracy:

> [I]t is (presumably) epistemically better to have higher credences in truths and lower credences in falsehoods. According to the cognitive decision-theoretic approach, epistemic rationality consists in taking steps that can reasonably be expected to <u>bring about epistemically good outcomes</u>. (Greaves and Wallace, 2006, 610)

Similarly for Gibbard:

> When a person forms her credences with epistemic rationality, our hypothesis will now run, it is as if she were voluntarily choosing her credences with the pure aim of truth—that is to say, to <u>maximize the expected accuracy of her credence</u>. (Gibbard, 2007, 149)

[31] A possible exception, within a different literature, appears in (Briggs, 2009). In a discussion of Dutch book arguments for van Fraassen's (1984) Reflection principle, Briggs suggests that "winning" and "losing" a bet on a proposition can be given a broader definition—winning if the proposition is true, losing otherwise—that doesn't require anyone taking the bet. Bets, on this understanding, are analogous to nonconsequentialist credal "acts." Briggs argues a dutch book only reveals incoherence if, at every possible world, the buyer of those bets "loses" more than he or she "wins"—independent of whether anyone in fact buys the bets in each world.

Leitgeb & Pettigrew are most naturally interpreted as endorsing a norm of having true beliefs or having credences that approximate the truth in the worlds where you have those beliefs:

> An epistemic agent ought to approximate the truth. In other words, she ought to minimize her inaccuracy. (Leitgeb and Pettigrew, 2010a, 202)
>
> It is often said that the epistemic norms governing full beliefs are justified by the more fundamental epistemic norm *Try to believe truths*... We will appeal to the more fundamental norm *Approximate the truth,* which is plausibly the analogue of the fundamental norm for full beliefs stated above. (Leitgeb and Pettigrew, 2010b, 236–7)

What these quotations have in common is that they can only naturally be interpreted as saying that rational beliefs should be formed with the aim of accuracy. But nonconsequentialist rules do not vindicate this intuition. As we've seen, nonconsequentialist rules will sometimes recommend performing credal acts that a rational agent can be *certain* will leave her farther from the truth than some alternative.

To summarize: nonconsequentialist rules don't seem to be the kinds of things that can apply to a person's credences. So they stand in need of a philosophical interpretation. Without one, they can't do the work they've been thought to do in epistemology. While it's at least prima facie intuitive that having credences that maximize expected accuracy, or are not accuracy-dominated, might be a good thing to do epistemically, it's not at all clear why following the nonconsequentialist rules is a good thing. So until these rules are given their own justification, they can't be used to justify probabilism or conditionalization.

What we can say is this: if epistemic utility theory is meant to bring decision theory to epistemology, then we should be concerned with the epistemic value people's credences could have if they take particular credal acts. But the nonconsequentialist rules aren't concerned with people's credences or credal acts. They are concerned with credence functions conceived as abstract objects: functions that exist at every possible world, whether or not they characterize anyone's doxastic state at any given world.

6 Logic and dominance

A natural hypothesis about why we should prefer N-dominance avoidance over C-dominance avoidance: in examples like #1 and #2, certain credences are closer to the truth at all *epistemically* possible worlds. But N-dominance avoidance only recommends credences if they are closer to the truth at all *logically* possible worlds. So its constraint is minimal in the extreme: it only prohibits acts that are worse no matter what the world is like, as a matter of pure logical necessity. And so (the thought goes), we can only question its normative force to the extent that we question the normativity of logic in reasoning.[32,33]

The problem with this response is that the difference between C-dominance avoidance and N-dominance avoidance is not merely a matter of epistemic vs. logical necessity. There are some ways in which our beliefs are not independent of their own accuracy as a matter of logical necessity.

Suppose at t you're uncertain what credence you'll have at t'. Your prior credence that you'll have credence n in H at t' is .5. Consider the utility of adopting, at t', various pairs of a first-order credence n in H and a second-order credence about whether you have credence n in H. Let A be the proposition that $Cr(H) = n$ at t'. Let's compare four such options:

	A	$\overline{A}$
a_1: $A \wedge Cr(A) = 1$	w_1	/
a_2: $\overline{A} \wedge Cr(A) = 0$	/	w_2
a_3: $A \wedge Cr(A) = .5$	w_3	/
a_4: $\overline{A} \wedge Cr(A) = .5$	/	w_4

It's clear that taking a_1 or a_2—the options where your second-order credence is correct—involves adopting more accurate credences than a_3 and a_4. Whichever of the first two options you take, it will certainly, predictably involve adopting a higher-order credence that is maximally accurate. So we might expect that, relative to these four options, a_3 and a_4 would be prohibited by both dominance rules and expected utility rules. But this turns out to be wrong in both cases when we supply nonconsequentialist versions of the rules.

First, consider nonconsequentialist expected accuracy: you can know in advance that either option a_1 or a_2 will lead to greater accuracy. So they should also have greatest expected accuracy. But a_3 and a_4 will have higher nonconsequentialist expected accuracy, relative to your prior.[34] From a decision-theoretic perspective, this is a bizarre result. It's not particularly strange to think that it might sometimes be rational to be uncertain about

[32]Thanks to Brandon Fitelson and Kenny Easwaran and (independently) Jason Turner and Robbie Williams for pressing me on this point.

[33]For agents in possession of knowledge, there are some logically possible worlds that aren't epistemically possible. What about epistemically possible worlds that aren't logically possible? This view doesn't presuppose that there are epistemically but not logically possible worlds—only that dominance reasoning in epistemology should consider all and only logically possible worlds. After all, suppose there's a logically impossible world where p and $\neg p$ are both true. Any credence that satisfies additivity will fail to be maximally accurate at this world: the maximally accurate credence function will assign 1 to p and to $\neg p$. Indeed, no credence functions, probabilistic or otherwise, would be dominated; the non-dominance constraint would be trivial.

[34] The partition of states that NEA uses is $\{w_1, w_2, w_3, w_4\}$. Let's assume your prior is maximally uncertain about which you'll perform. Then $NEA(a_1) = NEA(a_2) = \sum_w -.25(v_w(A) - Cr_{a_1}(A))^2 = -.5$ and $NEA(a_3) = NEA(a_4) = \sum_w -.25(v_w(A) - Cr_{a_3}(A))^2 = -.25$.

your future beliefs. But it is strange to think that *on the supposition that A is true*, a .5 credence in A has a higher expected accuracy than credence 1.

Second, consider N-dominance. While a_1 and a_2 strictly C-dominate a_3 and a_4, they don't N-dominate. After all, a_1 is closer than a_3 to the truth of w_1—but a_3 is closer at w_2. And so there are no N-dominated options.

If you perform one of the latter two options, you can be certain of at least one particular alternative option that it will land you greater accuracy. Whether the accuracy of the first-order credence is maximal or not, one of the first two options is certain to have greater global accuracy than all of the rest of the options. After all, they guarantee you a free true belief.

This illustrates a general point: the consequentialist rules will always require credences to be "transparent," in the following sense:

Transparency: $Cr(H) = n$ only if $[Cr(Cr(H) = n) = 1$ and for all $m \neq n$, $Cr(Cr(H) = m) = 0]$.

And this shows that there are some dependency relations between credal acts and their own accuracy that are logically necessary. It is logically necessary that any non-transparent credence function will be farther from the truth than a transparent alternative.

But N-dominance avoidance, like NEA maximization, permits non-transparent credence functions. A peculiar consequence: in examples of this form, N-dominance avoidance permits (and nonconsequentialist expected accuracy maximization requires) having Moore credences. By "Moore credences," I mean credences such that it would be natural to say of the agent that they believe H and believe that they don't believe H. Without taking a stand on the relation between belief and credence, we can offer $Cr(H) = 1$ and $Cr(\text{I don't believe } H) = 1$ as a natural candidate for such credences. If the person asserted both beliefs as a conjunction, they would assert a Moore sentence: "H but I don't believe H."

It's clear of any credal act that includes Moore credences that it is less accurate than a transparent alternative with the same first-order credences. While the person who adopts Moore credences may not be in a position to realize it, this fact is clear from a third-person perspective: an assessor who has precisely the same information as the agent about both H and whether the agent believes H will see that a person with Moore credences will be necessarily farther from the truth than a person with transparent credences.

The fact that sometimes N-dominance avoidance permits (and maximizing NEA requires) adopting Moore credences is an unexpected result. It reveals just how different nonconsequentialist rules are from their familiar consequentialist counterparts. Part of the justification that Joyce (2009) gives for why our credences should maximize expected accuracy by their own lights involves an appeal to the irrationality of having Moore credences:

> If, relative to a person's own credences, some alternative system of beliefs has a lower expected epistemic disutility, then, by her own estimation, that system is preferable from the epistemic perspective...This is a probabilistic version of Moore's paradox. Just as a rational person cannot fully believe "X but I don't

> believe X," so a person cannot rationally hold a set of credences that require her to estimate that some other set has higher epistemic utility. (277)

Joyce does not acknowledge that it's perfectly possible to have probabilistically coherent Moore credences, and such credence will maximize NEA from their own perspective.[35]

The question is: why do maximizing NEA and being non-N-dominated permit non-transparent credences, when transparent credences are *necessarily* closer to the truth? How can we justify N-dominance as an epistemic rule, given that it permits credences that are C-dominated as a matter of logical necessity?

At this point, the friend of N-dominance avoidance might concede that C-dominance avoidance has the same justification as N-dominance avoidance, as long as its state space covers all of logical space. That justification was that a credal act can't be rational if it is farther from the truth than some relevant alternative as a matter of logical necessity. Here, one might simply accept C-dominance avoidance.

But note that accepting even this extremely weak form of dominance avoidance does have some surprising implications. First, it requires us to give up subjective Bayesianism, the view that no probabilistic credences are rationally impermissible independent of evidence. We've seen that C-dominance avoidance prohibits probabilistic non-transparent credences.

Second, this places a big onus on the proponent of this weak version of dominance avoidance to explain why it only rules out credences that are dominated as a matter of logical necessity and not as a matter of epistemic necessity.

Finally, adopting this weak form of dominance avoidance means that no one can ever rationally doubt their own lower-order credences. This is a surprising result. Psychology is empirical: we are not rationally required to reject with absolute certainty the empirical hypothesis that our introspective knowledge of our own lower-order credences is fallible. If so, it must be possible to get evidence that rationally requires being less than maximally confident of our lower-order credences. In other words, it's possible to rationally have misleading evidence about one's credences, and it's rationally permissible to have positive credence that one has misleading evidence about one's credences.

7 Conclusion

I began by claiming that epistemic utility theory faces a dilemma. We have to choose between traditional consequentialist rules and nonconsequentialist rules. There are costs and benefits associated with each horn.

The benefit of the consequentialist rules is that they vindicate the idea that rational belief has the aim of accuracy, and so have an intuitive rationale. Their cost is that we'll have to give up probabilism, conditionalization, and other plausible epistemic norms.

[35] Assuming, as Joyce does, that we use a proper scoring rule. Proper scoring rules are accuracy measures according to which every probabilistic credence function maximizes NEA from its own perspective.

The benefit of nonconsequentialist rules is that we can retain familiar coherence constraints on credences: probabilism, conditionalization, etc. This conservatism comes with various perks: for example, it preserves the idea that Dutch-bookability indicates irrationality.[36] Moreover, the prescriptions of nonconsequentialist rules don't conflict with plausible evidential norms in the way that consequentialist rules do. The cost of nonconsequentialist rules, however, is that we'll have to give up on the idea that epistemic rationality is a matter of pursuing the goal of accuracy—and thereby give up our intuitive explanation for why we should obey these rules.

If we want to preserve Bayesian coherence norms, the second horn is more promising. But it leaves big unopened questions at the foundations of epistemic utility theory: why should we be concerned with the closeness of an abstract object to the truth, but *not* with the closeness of our own beliefs to the truth? If truth is the sole desideratum, then why not use consequentialist rules? What does it mean to assign mathematical objects epistemic value? Why are worlds where no one takes a certain credal act relevant to the assessment of that credal act?

Until these questions are answered, the nonconsequentialist rules stand in need of a philosophical interpretation. Before we can claim that epistemic utility theory justifies epistemic norms like conditionalization or probabilism, we needs to explain and justify the nonconsequentialist rules we're treating as fundamental.

Do the nonconsequentialist rules respect the norm of Evidence, instead of Accuracy? If they did, it would provide a special vindication of the research program of epistemic utility theory. Epistemic utility theory would be able to provide something that other forms of epistemic consequentialism so far haven't succeeded in providing: a mathematically precise epistemological theory that codifies the fundamental epistemic norm that we should believe what our evidence supports. The problem is that the arguments that philosophers in epistemic utility theory have provided for nonconsequentialist rules are stated in terms of the consequentialist goal of accuracy promotion. Our nonconsequentialist epistemic rules make recommendations that apparently align with evidentialism. Now we just need to find out *how* and *why* they do.[37]

---

[36]However, this form of practical irrationality need not *explain* why violating the familiar coherence constraints is epistemically irrational.

[37]Many thanks to Rachael Briggs, Michael Caie, Catrin Campbell-Moore, Kenny Easwaran, Branden Fitelson, Hilary Greaves, Sophie Horowitz, Richard Pettigrew, Bernhard Salow, Miriam Schoenfield, Jason Turner, Ian Wells, Robbie Williams, and especially Ryan Doody. Thanks also to audiences and organizers of the Epistemic Consequentialism Conference, Konstanz (2015), the Choice Group, LSE (2015), the Eastern APA Conference, Philadelphia (2014), the Epistemic Consequentialism Workshop, LSE (2014), the Formal Epistemology Workshop, USC (2014), and the Epistemic Utility Theory Conference, University of Bristol (2013), as well as audiences at UIUC, USC, and Yale. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 312938'.

References

Audi, R. (1973). Intending. *Journal of Philosophy*, 70(13):387–403.

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*.

Briggs, R. (2009). Distorted reflection. *Philosophical Review*, 118(1):59–85.

Caie, M. (2012). Belief and indeterminacy. *Philosophical Review*, 121(1):1–54.

Caie, M. (2013). Rational probabilistic incoherence. *Philosophical Review*, 122(4):527–575.

Christensen, D. (1996). Dutch-book arguments depragmatized: Epistemic consistency for partial believers. 93(9):450–479.

Davis, W. A. (1984). A causal theory of intending. *American Philosophical Quarterly*, 21(1):43–54.

Easwaran, K. and Fitelson, B. (2012). An ‘evidentialist’ worry about joyce’s argument for probabilism. *Dialectica*, 66(3):425–433.

Feldman, R. and Conee, E. (1985). Evidentialism. *Philosophical Studies*, 48(1):15–34.

Gibbard, A. (2003). Thoughts and norms. *Philosophical Issues*, 13(1):83–98.

Gibbard, A. (2005). Truth and correct belief. *Philosophical Issues*, 15(1):338–350.

Gibbard, A. (2007). Rational credence and the value of truth. In *Oxford Studies in Epistemology*, volume 2, pages 143–164. Oxford University Press.

Goldman, A. (1979). “what is justified belief?”. In Pappas, G., editor, *Justification and Knowledge*, pages 1–25. Boston: D. Reidel.

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

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

Harman, G. (1976/1997). Practical reasoning. In Mele, A. R., editor, *Review of Metaphysics*, pages 431–63. Oxford University Press.

James, W. (1896). The sentiment of rationality. In *The Will to Believe and Other Essays in Popular Philosophy*. Synthèse, Norwood, MA.

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 *Degrees of Belief*, volume 342, pages 263–297. Synthèse.

Joyce, J. (2013). Accuracy, self-accuracy, and choice. Presented at the Bristol Epistemic Utility Theory Conference.

Joyce, J. (ms). Why evidentialists need not worry about the accuracy argument for probabilism.

Kelly, T. (2006/2014). Evidence. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*.

Kennedy, R. and Chihara, C. (1979). The dutch book argument: Its logical flaws, its subjective sources. *Philosophical Studies*, 36(1):19–33.

Kim, J. (1988). What is "naturalized epistemology?". *Philosophical Perspectives*, 2:381–405.

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

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

Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, 59:5–30.

Pettigrew, R. (2013). Accuracy and evidence. *Dialectica*, 67(4):579–596.

Rosenkrantz, R. D. (1981). *Foundations and Applications of Inductive Probability*. Ridgeview Press, Atascadero, CA.

Ross, J. (2009). How to be a cognitivist about practical reason. *Oxford Studies in Metaethics*, 4:243–281.

Shah, N. (2003). How truth governs belief. *Philosophical Review*, 112(4):447–482.

Sinhababu, N. (2013). The desirebelief account of intention explains everything. *Noûs*, 47(4):680–696.

van Fraassen, B. C. (1984). Belief and the will. *Journal of Philosophy*, 81(5):235–256.

Velleman, D. (1989). *Practical Reflection*. Princeton University Press, Princeton, NJ.

Velleman, D. (2000). On the aim of belief. In *The Possibility of Practical Reason*, pages 244–281. Oxford University Press.

Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, 16(s16):267–97.

Wedgwood, R. (ms). The pitfalls of 'evidence'.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.