# Epistemic Utility Theory and the Aim of Belief

It's widely accepted that rational belief aims at truth.[1] Objectively correct belief is true belief. A more controversial question: 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 for rational belief to have the aim of approximating the truth. By combining the tools of decision theory with an epistemic form of value—gradational accuracy, proximity to the truth—we can justify various epistemological norms. These arguments generally use one of two tools: the notion of *expected accuracy* and 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 shift our beliefs in response to new evidence in a certain way—by a method called "conditionalization"—is that doing so uniquely maximizes expected accuracy (Greaves & Wallace 2006; Leitgeb & Pettigrew 2010b).

I'm going to argue that deriving these results requires using 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 coherence norms that epistemic utility theory had hoped to justify. Indeed, the prescriptions of epistemic utility theory conflict with these norms.

Furthermore: the things epistemic utility theorists 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 *this thing*—whatever it is—doesn't explain why we should conditionalize on our evidence.

---

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

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 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 **observational 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 observational rules, we can avoid predicting rational violations of probabilistic coherence, conditionalization, and so on—but at the cost of giving up the idea that epistemic rationality is a matter of having those credences that best approximate the truth. This means giving up our only 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 observational 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 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 "observational expected accuracy" or "OEA." I argue that closely related problems afflict accuracy "dominance" arguments: traditional (consequentialist) Dominance 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 "observational dominance" or "O-dominance." The prohibition of O-dominated credences doesn't lead to conflicts with epistemic norms like conditionalization.

In section 4, I show that traditional consequentialist rules conflict not only with

[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.

conditionalization but also with probabilism. While some have argued on these grounds that it can be rational to have probabilistically incoherent credences, I argue 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 observational rules can generate better results than the more familiar consequentialist rules (in that they preserve conditionalization and probabilism), they have 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 observational rules, as we could with consequentialist rules.

In section 6, I consider and reject the argument that O-Dominance derives its plausibility immediately from the normativity of logic.

In section 7, I consider the bigger picture. The reason why the familiar consequentialist rules generate conflicts with conditionalization and probabilism is that they conflate the epistemological ideal of respecting the evidence with the ideal of having true belief. Respecting the evidence involves an asymmetric relation between belief and world: it has a mind-to-world direction of fit. But having true beliefs is symmetric: there's no one-way direction of fit. You can attain that goal equally well by fitting the world to your beliefs.

I argue that a good philosophical interpretation of observational rules would be one that made explicit the relation between OEA-maximal or non-O-dominated belief and evidential support. If we want to answer the foundational questions about epistemic utility theory, this is where we should start.

# 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]

---

[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:* $Cr(\mathscr{W}) = 1$

**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 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.

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,[5] 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 (i) if another is more accurate at every possible world and (ii) there is some credence function that is not dominated.[6] Joyce then proves that for a range of epistemic utility functions satisfying certain constraints, any credence function that violates probabilism is strongly dominated.

Instead of Dominance, Greaves & Wallace (2006) and Leitgeb & Pettigrew (2010a,b) argue for coherence norms by appeal to the decision rule Maximize Expected (Epistemic) Utility. Expected utility is the average utility of a set of possible outcomes, weighted by some relevant measure of probability. Greaves & Wallace argue that, upon receiving new evidence, the update procedure that uniquely max-

---

3. *Finite additivity:* if $X$ and $Y$ are disjoint, then $Cr(X \vee Y) = Cr(X) + Cr(Y)$

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

[5] 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.

[6] I'll call the decision rule "Dominance" with a capital D and the relation "dominance" with a lower case d. In what follows, I will generally leave off the second condition.

imizes expected accuracy is to conditionalize 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 (namely, 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 from those used in practical decision theory. So far, for the most part, the difference hasn't been acknowledged or justified. At most, the rules of epistemic utility theory are understood as minor simplifications of the real decision-theoretic rules. As I'll show, the differences aren't innocuous. First, they produce very different results; different epistemic actions are sanctioned. Second, they require a very different philosophical interpretation: they cannot be applied to a person's credal acts or credal states. Without the intuitive motivations of the old rules, the new rules have no clear motivation at all.

## 2 How epistemic acts affect their own accuracy

### 2.1 The central example

Greaves & Wallace (2006) and Leitgeb & Pettigrew (2010b) argue that the rule Maximize Expected Accuracy always requires agents to conditionalize on any new evidence:

**Maximize Expected Accuracy (MaxExAcc):** rational agents have the credences that maximize expected accuracy (from the agent's perspective).

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

The kind of cases where MaxExAcc 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

[7] See Berker (forthcoming), (Caie forthcoming), (Greaves forthcoming).

might be self-verifying,[8] self-falsifying,[9] etc. Your credence in a proposition can also causally affect the 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

---

[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).

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

[11] Typically you believe you'll do what you intend to do; this belief is, on some views, constitutive of the intention.

[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 forthcoming).

[13] Examples of the first and second kind are central to Berker's (2013) argument against various forms of epistemic consequentialism, such as reliabilism. 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 it epistemically required for her 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

that proposition to be more or less likely to be true. I'll return to this example throughout this paper.

**Example #1**

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 $[0,1]$, $Cr_1(H) = n$ at $t_1$ will make it the case that $Ch_2(H) = n$ 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 a world and the chances 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 MaxExAcc, 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 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 MaxExAcc and Dominance say? We'll turn to MaxExAcc first and return to Dominance in section 3.2.

---

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.

[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. There are (slightly less realistic) examples without this confounding factor: for example, suppose your beliefs determine whether it'll rain on some uninhabited planet.

## 2.2 Two kinds of decision theory

In order to know what MaxExAcc recommends, we need to specify what the rule says. MaxExAcc is a special case of a decision-theoretic rule, Maximize Expected Utility, paired with an accuracy-based utility function. There are two competing forms that the rule Maximize Expected Utility most commonly takes in decision theory: Maximize Causal Expected Utility and Maximize Evidential Expected Utility. As we will see, both of these rules are different in interesting ways from the observational 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 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 decision 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[15]:

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

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 $Cr_b$:[16]

$$\begin{array}{llll} k_1: & \text{you adopt } Cr_a \;\Box\!\!\rightarrow H & k_2: & \text{you adopt } Cr_a \;\Box\!\!\rightarrow H \\ & \text{you adopt } Cr_b \;\Box\!\!\rightarrow H & & \text{you adopt } Cr_b \;\Box\!\!\rightarrow \overline{H} \\ & \text{you adopt } Cr_c \;\Box\!\!\rightarrow \overline{H} & & \text{you adopt } Cr_c \;\Box\!\!\rightarrow \overline{H} \end{array}$$

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. As an accuracy measure, we'll just use the negative Brier score.[17] Let $v_w(\cdot)$ be the omniscient probability function at a world $w$ (map-

[15] 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.

[16] 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. I prefer the non-Lewisian variation because it allows full specification of the values of outcomes even if they're chancy. We can't say how accurate a credence in $H$ is at a state unless we specify whether $H$ is true at that state.

[17] The Brier score is widely accepted as a good inaccuracy measure; see (Joyce 2009) and (Leitgeb & Pettigrew 2010a) for philosophical defenses of it. It's negative because I've been talking about maximizing accuracy, rather than minimizing inaccuracy. Ideally accurate credences have a Brier score of 0; nonideal credences have a negative accuracy score.

ping all truths to 1 and falsehoods to 0). Then the negative Brier score of $Cr(X)$ at $w$ is:

$$U(Cr(X),w) = -(v_w(X) - Cr(X))^2$$

Now we can assign values to the worlds in our decision problem:

| | $k_1$ | $k_2$ |
|---|---|---|
| $Cr_a$ | 0 | 0 |
| $Cr_b$ | $-.25$ | $-.25$ |
| $Cr_c$ | 0 | 0 |

There's no need to calculate the causal expected accuracy because $Cr_b$ is strongly dominated.[18] So we know that $Cr_b$ is impermissible, according to Maximize CEA.

Maximize Causal Expected Accuracy (CEA) therefore conflicts with conditionalization. Conditionalization requires you to adopt $Cr_b$, whereas Maximize CEA requires you to adopt either $Cr_a$ or $Cr_c$.

### 2.4 Evidential expected accuracy

Now we'll figure out what Maximize Evidential Expected Accuracy (EEA) recommends. First, note that only four possibilities ($w_1, w_2, w_3, w_4$) have positive probability.

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

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

[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.

The evidential expected accuracy calculation is as follows. 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]

$$EEA(Cr_a(H)) = Cr_0(\mathsf{w_1} \mid \mathsf{Cr_a})(-(v_{w_1}(H) - Cr_a(H))^2) = 0$$
$$EEA(Cr_b(H)) = Cr_0(\mathsf{w_2} \mid \mathsf{Cr_b})(-(v_{w_2}(H) - Cr_b(H))^2) + Cr_0(\mathsf{w_3} \mid \mathsf{Cr_b})(-(v_{w_3}(H) - Cr_b(H))^2) = -.25$$
$$EEA(Cr_c(H)) = Cr_0(\mathsf{w_4} \mid \mathsf{Cr_a})(-(v_{w_4}(H) - Cr_a(H))^2) = 0$$

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 measures the distance between credences and truth. If your credence in $H$ is extremal, the distance between your credence and the truth will be 0. But for any non-extremal credence, there will be some distance between your credence and the truth.

So no matter what your prior is, Maximize EEA instructs you to adopt extremal credences. And this means that, like Maximize CEA, 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 & Wallace (2006) and Leitgeb & Pettigrew (2010b) argued that updating by conditionalization uniquely maximizes expected accuracy, from the perspective of any probabilistic priors. I've shown that if MaxExAcc 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. And there is a fall-back position that allows one to do so in the example I gave: by ruling nonextremal credences in $H$ irrational. We can't get the right results for *all* probabilistic credences, one might say; but we can at least get it for the rational ones.

Here are three reasons for skepticism about this strategy. First, it's intuitively 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

[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).

like this; and so there's no telling whether $H$ will be true." 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. So why would it be irrational to have a credence like .5, which is also symmetrical about whether $H$ is true? Third, what's rational to believe depends on what the evidence supports. But causal and evidential MaxExAcc only permit credence 1 or 0 in $H$. 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![20]

The better conclusion to draw from these sorts of examples is 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 7.[21]

# 3 Observational rules

## 3.1 Consequentialist and observational expected accuracy

Now, you might think I have provided a counterexample to the proofs in Greaves & Wallace (2006) and Leitgeb & Pettigrew (2010b) that MaxExAcc entails 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:** $CEA_{Cr}(Cr') = \sum_{k \in K} Cr(k) U(Cr', k)$

[20] A natural thought: 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 $H$. 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.

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

Call the decision rules Maximize CEA and Maximize EEA **consequentialist** rules: they are concerned with the outcomes of credal acts. Philosophers working on epistemic utility theory, including Greaves & Wallace and Leitgeb & Pettigrew, use a notion of expected accuracy different from either causal or evidential expected accuracy. We'll call their inaccuracy measure "observational expected accuracy" or "OEA": it measures expected accuracy for pure observers, whose epistemic acts have no dependency relations with facts about the world, roughly spaking.[22] Where $E$ is the set of epistemically possible worlds,

**Observational expected accuracy:** $OEA_{Cr}(Cr') = \sum_{w \in E} Cr(\mathrm{w})U(Cr', w)$

The rule Maximize OEA does recommend the same epistemic acts as conditionalization.

How? OEA 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 your performing that epistemic act.* Causal expected accuracy uses partitions of worlds that are based on the effects your acts might have. In both cases, the weighting on the utilities for the expectation depends on your taking the act in the worlds considered.

### 3.2 Consequentialist and observational accuracy dominance

There is a similar distinction to be drawn for the decision rule Dominance: it comes in consequentialist and observational 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 "observational dominance" or "O-dominance": it is dominance for a pure observer, whose credal acts have no effect on the world. Here's how Joyce defined the relation of O-dominance:

> $Cr$ strongly **[O-]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 definition of dominance in practical decision theory:

---

[22] As we'll see below, even this characterization of OEA is an oversimplification. It is tricky to find an intuitive characterization of what exactly OEA is.

> $Cr$ strongly **dominates** $Cr'$ relative to $U$ iff $U(\mathsf{Cr} \cap s) > U(\mathsf{Cr} \cap s)$ for all $s \in S$.[23]

The difference: O-dominance, unlike dominance, takes into account the value of acts at *all* worlds.

If we use the decision-theoretic conception of dominance for the purposes of epistemic utility theory, 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-dominated are irrational.

Consider, again, the self-fulfilling example. Since all of the credence functions we considered were probabilistically coherent, none were O-dominated.[24] But $Cr_b$—along with any other non-extremal credence in $H$—will be dominated. And so they'll be ruled epistemically impermissible. So if we move to the consequentialist rule Dominance, we will find conflicts with conditionalization. Credence functions that seem perfectly reasonable will be ruled rationally impermissible.

The difference between observational rules and traditional, consequentialist rules has gone largely unnoticed. Those who have noticed this difference (Greaves (forthcoming) and Caie (forthcoming)) have treated the observational rules as convenient simplifications, not competitors to consequentialist rules. They have argued that we should "fix" epistemic utility theory by replacing observational 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 make the world match itself. And so, as I'll argue in much of the rest of the paper, OEA's failure to consider the consequences of belief is a feature, not a bug. [25]

---

[23] 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 aren't 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.

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

[25] A switch from OEA to CEA or EEA would have wide-ranging effects on epistemic utility theory: not only could we not get the same conclusions, but we wouldn't even be able to rely on the same premises. A standard starting point in epistemic utility theory is to characterize plausible epistemic utility functions. A quite weak constraint on epistemic utility functions, which Gibbard (2007) introduced and Joyce and Leitgeb & Pettigrew defend, is that epistemic utility functions must make rational credences *immodest,* in the following sense:

**Immodesty:** Any rational credence $Cr$ will assign itself higher expected epistemic utility than any other credence function.

# 4 Coherence norms or consequentialist rules?

## 4.1 Probabilism, dominance, and expected accuracy

Caie uses an accuracy dominance argument to defend rational violations of probabilism. His preferred examples involve self-referential beliefs, but can be easily generalized to causal examples.

> **Example #2**
>
> 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 $H$:
>
> $$H \text{ is true iff } Cr_1(H) \leq .5$$
>
> At $t_0$, what credence should you have in $H$, and in not-$H$, conditional on $H \leftrightarrow Cr_1(H) \leq .5$?

---

As Joyce notes, a variety of different epistemic utility functions satisfy this desideratum. (These epistemic utility functions are called "strictly proper" scoring rules.) This is so, though, only with OEA. What happens if we use CEA or EEA?

The handstand example is an obvious case where the Brier score is compatible with modest probabilistically coherent credences when expected accuracy is interpreted causally. (I'll leave it an exercise to the interested reader to see why; simply consider the perspective of $Cr_b$.) The argument involving EEA is more complicated. We can call a credence function "transparent" iff it conforms to this:

**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 global 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 of your own credence function won't be higher than any other credence function: all alternatives will have undefined EEA.

So the Brier score is not a proper scoring rule when expected accuracy is interpreted as CEA or EEA! And in fact, any plausible accuracy measure will also be improper: extremal credences in $H$ will perfectly match the truth, and so their local inaccuracy will be 0 by any measure of distance from the truth. By contrast, nonextremal credences won't *perfectly* match the truth, and so their inaccuracy will be non-zero by any plausible inaccuracy measure. If we want to use an Immodesty constraint on epistemic utility functions, we need to do so with OEA. And there is much to be said for Immodesty: in particular, that, combined with Joyce's (2009) argument that nonprobabilistic credences are necessarily accuracy dominated, it entails that all and only probabilistic credences are non-dominated.

The credence in $H$ that is non-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 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 (Dominance, Maximize CEA, Maximize EEA) will sometimes recommend violating probabilism. Caie (forthcoming) 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.

I claimed in the introduction that epistemic utility theory faces a dilemma: we must adopt a consequentialist decision theory that reflects how actions can affect the world (thereby giving up conditionalization and probabilism), or else give an non-consequentialist justification for an observational decision theory (thereby giving up the idea that rational belief aims for accuracy). To say that maximizing OEA or avoiding O-domination ensures that our beliefs are closer to the truth can't be right. Examples #1 and #2 are counterexamples: in both cases a rational agent can be certain that credences that violate Maximize OEA and O-Dominance are closer to the truth.

I interpret Caie and Greaves as accepting the first horn of the dilemma. If rationality requires aiming for true belief, then we should abandon observational rules. But I tollens where they ponens. I think examples like #1 and #2 show that rationality doesn't require us to aim for true belief.

Focusing on a self-frustrating case like example #2, as Caie does, is misleading: 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.

## 4.2 Can an act/state distinction explain away the worry?

Joyce (2013) argues that cases like example #2 can be explained away without fanfare. 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 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 O-dominated.[26]

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 (an option 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 gives maximal 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 Dominance and

[26] Neither Joyce nor Caie discusses the distinction between dominance and O-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 dominates all alternatives; but it is O-dominated. That is, there is a probabilistic credence function $Cr_p$ which O-dominates $Cr_n$—but only if the relevant agent doesn't adopt $Cr_p$.

Maximize CEA/EEA. If we give up the latter, 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 you don't take them

The difference between consequentialist rules and observational 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 you adopt it. Observational 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 relevant agent doesn't take that act.

This fact is puzzling. What is the value of a credal act at an outcome where the act isn't taken? 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 adopting that credence function* at that world, or the *accuracy of your doxastic state* at that world.[27] And so we cannot interpret observational rules as means to an epistemic end.

Consider Joyce's (1998, 2009) "dominance" argument for probabilism. The argument is, of course, actually an O-dominance argument for probabilism. A premise of his argument is that it is irrational to adopt credences that are accuracy O-dominated. But the rationale for that premise equally motivates the claim that it's irrational to adopt credences that are accuracy dominated. Here is Joyce's (2009) justification for the claim that it's irrational to adopt O-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 O-Dominance 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

[27] Note that causal decision theory (but not evidential decision theory) allows you to calculate the *expected* value of an act, given that you 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 you don't take it.

the defect is one of epistemic rationality" justifies Dominance at least as well as O-Dominance. After all, adopting 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 O-dominated is at the very least substantive.

So: we need some justification for the claim that choosing options that are O-dominated would be irrational, given that choosing options that are dominated is not. And similarly, we need some justification for the claim that choosing options that are not OEA-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 credences*?

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

Here is a natural defense one might give of observational rules. Consequentialist rules imported from decision theory generate conflicts with conditionalization and probabilism precisely *because* they reflect the causal and evidential dependency relations between our credences and the world. What we should aspire to believe, this argument goes, is what's true *independent of our beliefs*. And observational rules allow us to do so. For example, observational expected accuracy allows us to measure the expected proximity of our belief states to this independent truth.

The problem with this response is that there are a variety of ways in which, intuitively, our beliefs do affect what is *rational* for us to believe, and not just what's true. For example, if I know $Cr(H) = Ch(H)$ and I know $Cr(H) = .4$, then I should infer $Ch(H) = .4$. But that fact about chance isn't independent of my beliefs. If I know that my friend tries to learn from me by acquiring some of my beliefs and I know that he's reasonably reliable at determining what I believe (perhaps by conversing with me), then I know that whether he believes $H$ isn't entirely independent of my beliefs. And so on.

## 5.2 Credal acts and observational rules

I'm going to show that observational 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.

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

Let's return once more to the self-fulfiling handstand example. You know in advance that if you adopt $Cr(H) = 1$, the (local) accuracy your credences will have is 0 (i.e. perfect; its distance from the truth is 0). But the (local) OEA of credence 1 in $H$, from your point of view, will not be 0. The only assumption we need to make about the accuracy measure is that a credence in a proposition matches the proposition's truth value iff its accuracy is 0. Then, in examples like this, the known accuracy of a credal act will not match its OEA.

So OEA cannot be the expected accuracy of a person's credences.

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

| | $s_1$ | $s_2$ |
|---|---|---|
| $a_1$ | $w_1$ | $w_2$ |
| $a_2$ | $w_3$ | $w_4$ |

In ordinary 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, $a_1$ dominates $a_2$. On the other hand, in O-dominance arguments, we assess whether the value of $a_1$ is greater than the value of $a_2$ at *all four outcomes*, including, e.g., $w_3$. 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 observational rules

Where the distinction between consequentialist and observational rules has been discussed, observational rules are treated as a naive oversimplification, to be abandoned for the more robust consequentialist rules whenever cases with more complex dependency relations arise. Observational rules have never been treated as competing with the consequentialist rules—as alternative rules to be considered on their own merits.

And so the intuitive justifications that have been given for the observational 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 observational rules don't always do that. In examples #1 and #2, observational 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 observational rules that actually support consequentialist rules, and in some cases are not even consistent with observational rules. I underline passages that reveal the consequentialist assumptions that conflict with observational 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 & Wallace's justification for maximizing accuracy is straightforwardly incompatible with maximizing observational 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 & 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)

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 <u>her inaccuracy</u>. (Leitgeb & Pettigrew 2010a, 202)
>
> It is often said that the epistemic norms governing full beliefs are justified by the more fundamental epistemic norm <u>*Try to believe truths*</u>. . . 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 & 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 observational rules do not vindicate this intuition. As we've seen, observational 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: observational rules don't seem to be the kinds of things that can apply to a person's credences or credal acts. 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 observational 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 observational 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 O-Dominance over Dominance: in examples like #1 and #2, certain credences are closer to the truth at all *epistemically* possible worlds. But O-Dominance 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.[28]

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

The problem with this response is that the difference between Dominance and O-Dominance 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 observational versions of the rules.

First, consider observational 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 observational expected accuracy, relative to your prior.[29] 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 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 O-dominance. While $a_1$ and $a_2$ strictly dominate $a_3$ and $a_4$,

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

they don't O-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 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 O-Dominance, like observational MaxExAc, permits non-transparent credence functions. A peculiar consequence: in examples of this form, O-Dominance permits (and observational MaxExAcc 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: "$H$ is true and 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 O-Dominance permits (and maximizing OEA requires) adopting Moore credences is an unexpected result. It reveals just how different observational 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 if you have probabilistically coherent Moore credences, then you will maximize OEA from your own perspective.[30] After all, there's no conflict between Moore credences and probabilism.

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

At this point, the friend of O-Dominance might concede that Dominance has the same justification as O-Dominance, 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 Dominance.

But note that accepting even this extremely weak form of Dominance 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 Dominance prohibits probabilistic non-transparent credences. Second, this places a big onus on the proponent of this weak version of Dominance 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. And finally, this means that no one can ever rationally doubt their own lower-order credences. This is a surprising result. Psychology is empirical and we are not rationally required to confidently reject the idea that our introspective knowledge of our own lower-order credences is infallible. And so intuitively it must be possible to get evidence that rationally requires doubting your lower-order credences.

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

## 7 Accuracy and evidence

I've argued that the notions of expected accuracy and accuracy dominance that are used in epistemic utility theory don't have a clear philosophical interpretation. And so some important questions about the foundations of epistemic utility theory remain open. By way of conclusion, let me discuss what I think are the main philosophical stakes in answering these questions.

Here are two claims that are pretty plausible:

**Evidence** You should adopt the belief state that your evidence supports.

**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 different 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.

But these claims are not the same. There's an intuitive sense in which, in cases like the self-fulfilling handstand example, evidential support and promotion of accuracy simply come apart. Accuracy demands extremal credences in that example. Intuitively, Evidence doesn't.[31]

What does Evidence require in that example? It's not at all clear. Here is a (non-exhaustive) list of some positions one might take with respect to the self-fulfilling handstand example:

1. **Impermissivism:** there's a particular credence in $H$ that's rationally required.
2. **Permissivism:** more than one credence in $H$ can be rational, as long as it's arrived at by rational update on permissible priors.

[31] 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 & Fitelson (2012) argue that (one form of) the Dominance principle interacts with the Principal Principle in a way that generates an undesirable order-dependence. Joyce (manuscript) and Pettigrew (forthcoming) argue that the worry is avoidable if we either switch to a weaker form of Dominance or to objective expected accuracy.

3. **Extreme permissivism:** more than one credence in $H$ is rational, and the choice is unconstrained by your priors.

4. **Gappy credences:** you shouldn't have any credence at all in $H$.[32]

5. **Epistemic dilemma:** there are no permissible doxastic states you can adopt.

6. **Belief indeterminacy:** you should adopt an indeterminate doxastic state.

7. **Normative indeterminacy:** it's indeterminate what doxastic state you should have.

For my part, I'm actually somewhat attracted to option 7. Your evidence can only support confidence in $H$ to the extent that it supports confidence that you'll believe $H$. After all, your credence in $H$ is the sole determinant of how likely $H$ is to be true. So if anyone wants to find out whether $H$, they must investigate how confident you'll be in $H$. But how do we find out what you'll believe? We've stipulated that you're rational (because we're trying to find out what attitude is rational to take, and we do so by considering what attitudes rational agents take). But wait: if you're rational, you'll be confident of $H$ only on the grounds that the evidence supports $H$. After all, rationality requires believing $H$ only to the extent that the evidence supports $H$.

So there is no independent evidential grip we can get on $H$. Before we can find out what credence is rational in $H$, we need to find out what credence you have in $H$ (since this determines whether $H$); but before we can find out what credence you have in $H$, we need to know what credence is rational in $H$ (because you are, ex hypothesi, rational). There doesn't seem to be an evidential basis for *any* credence here. So it seems plausible to me that the norm of Evidence doesn't give any verdict in this case whatsoever.[33]

---

[32] Perhaps because you shouldn't take any attitude toward the proposition that you adopt this or that credence in $H$. Some ((Levi 1997), (Spohn 1977), Briggs (personal communication)) say we shouldn't, or needn't, have credences in our own future acts—at least in the context of deliberation.

[33] Note that even if we add extra evidence—say, that a 80% of people who were in your position and learned $Cr(H) = Ch(H)$ ended up making $H$ true—I think it's rational for that evidence to be screened by $Cr(H) = Ch(H)$, in the sense of Weatherson (manuscript). That is:

1. $Cr(H \mid 80\% \text{ info}) > Cr(H)$
2. $Cr(H \mid 80\% \text{ info} \wedge Cr(H) = Ch(H)) = Cr(H \mid Cr(H) = Ch(H))$

In the light of learning $Cr(H) = Ch(H)$, the information about other people is neutralized. So it's not that you have *no* evidence about $H$. It's that what evidence you have doesn't seem to make any push toward any credence.

I began by claiming that epistemic utility theory faces a dilemma. We have to choose between traditional consequentialist rules and observational 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.

The benefit of observational rules is that we can retain probabilism, conditionalization, etc. Their cost 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.

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 observational 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 non-consequentialist rules we're treating as fundamental.

Do the observational rules respect the norm of Evidence, instead of Accuracy? If they do, 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 observational rules are stated in terms of the consequentialist goal of accuracy promotion. And so, if the rules associated with OEA and O-dominance make recommendations that align with evidentialism, it's not at all clear *how* or *why* they do. I think these are the questions we should be answering.

## References

Berker, Selim (2013). "Epistemic Teleology and the Separateness of Propositions." *Philosophical Review*, 122: pp. 337–393.

Berker, Selim (forthcoming). “The Rejection of Epistemic Consequentialism.” *Philosophical Issues*.

Caie, Michael (2012). “Belief and Indeterminacy.” *Philosophical Review*, 121(1): pp. 1–54.

Caie, Michael (forthcoming). “Rational Probabilistic Coherence.” *Philosophical Review*.

Easwaran, Kenny, and Branden Fitelson (2012). “An ‘Evidentialist’ Worry about Joyce’s Argument for Probabilism.” *Dialectica*, 66(3): pp. 425–433.

Firth, Roderick (1978/1999). “The Schneck Lectures.” In J. Troyer (ed.) *In Defense of Radical Empiricism: Essays and Lectures*, Lanham, MD: Rowman & Littlefield, pp. 315–67.

Gibbard, Allan (2003). “Thoughts and Norms.” *Philosophical Issues*, 13(1): pp. 83–98.

Gibbard, Allan (2005). “Truth and Correct Belief.” *Philosophical Issues*, 15(1): pp. 338–350.

Gibbard, Allan (2007). “Rational Credence and the Value of Truth.” In *Oxford Studies in Epistemology*, Oxford University Press, vol. 2, pp. 143–164.

Greaves, Hilary (forthcoming). “Epistemic Decision Theory.” *Mind*.

Greaves, Hilary, and David Wallace (2006). “Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility.” *Mind*, 115(459): pp. 607–632.

James, William (1896). “The Sentiment of Rationality.” In *The Will to Believe and Other Essays in Popular Philosophy*, Norwood, Mass.: Synthèse.

Joyce, James (1998). “A Nonpragmatic Vindication of Probabilism.” *Philosophy of Science*, 65(4): pp. 575–603.

Joyce, James (2009). “Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief.” In *Degrees of Belief*, Synthèse, vol. 342, pp. 263–297.

Joyce, James (2013). “Accuracy, Self-Accuracy, and Choice.” Presented at the Bristol Epistemic Utility Theory Conference.

Joyce, James (manuscript). “Why Evidentialists Need Not Worry About the Accuracy Argument for Probabilism.”

Leitgeb, Hannes, and Richard Pettigrew (2010a). "An Objective Justification of Bayesianism I: Measuring Inaccuracy." *Philosophy of Science*, 77(2): pp. 201–235. In Firth (1978/1999).

Leitgeb, Hannes, and Richard Pettigrew (2010b). "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy." *Philosophy of Science*, 77(2): pp. 236–272.

Levi, Isaac (1997). *The Covenant of Reason: Rationality and the Commitments of Thought*. Cambridge University Press.

Lewis, David (1981). "Causal Decision Theory." *Australasian Journal of Philosophy*, 59: pp. 5–30.

Pettigrew, Richard (forthcoming). "Accuracy and Evidence." *Dialectica*.

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

Spohn, Wolfgang (1977). "Where Luce and Krantz Do Really Generalize Savage's Decision Model." *Erkenntnis*, 11(1): pp. 113–134.

Velleman, David (2000). "On the Aim of Belief." In *The Possibility of Practical Reason*, Oxford University Press, pp. 244–281.

Weatherson, Brian (manuscript). "Do Judgments Screen Evidence?"

Wedgwood, Ralph (2002). "The Aim of Belief." *Philosophical Perspectives*, 16(s16): pp. 267–97.