Impractical Decision Theory

Gabrielle Kerbel[1] *& Adrian Liu,*[2] *April 1, 2025, NYU Tech Support*

[1] University of Michigan, ✉ gekerbel@umich.edu

[2] Rutgers University, ✉ adrian.liu@rutgers.edu

1 | Imptroduction.

Epistemic Decision Theory (EDT) evaluates epistemic states by how *accurate* they are.

1. Represent epistemic states with *credence functions* c which take propositions as input and outputs a value in [0,1] corresponding to the agent's confidence.[3]
2. Measure accuracy at a world w with a *scoring rule* that takes a credence function c and a world w and outputs a number $A\langle c, w\rangle$ representing c's accuracy at w.
3. Measure accuracy over a universe W via *expected* accuracy: $EA_\pi(c) =_{\text{df}} \sum_{w\in W} \pi(w)\cdot A\langle c, w\rangle$. We assume that scoring rules are *strictly proper* – that is, they satisfy $\forall \pi, c : EA_\pi(c) \leq EA_\pi(\pi)$, with equality if and only if $\pi = c$.[4]

[3] Propositions are modeled as sets of possible worlds $\{w_1, w_2, \ldots\}$ drawn from a universe W.

[4] Strict propriety guarantees that credence functions are *immodest* in that they do not consider any other credence function to be more accurate in expectation, and that the scoring rule is *truth-directed* in that at any world w, credence functions are more accurate when they move uniformly toward the truth values of propositions at w. These are arguably constitutive of the notions of accuracy and expectation, if anything is.

EDT's **Accuracy Norm**: rational agents maximize expected accuracy. *Formally*: Let $\phi[c']$ mean that the (relevant) agent adopts credence function c'. Then an agent with credence c is rational only if

$$c = \arg\max_{c'\in\mathbb{Cr}}[EA^{\mathbb{E}}_\pi(\phi[c'])],$$

where A is a scoring rule, $\mathbb{E}$ is an estimation method, and π describes the credence function relative to which estimation is made.
Three ways to select π:

$$c = \arg\max_{c'\in\mathbb{Cr}}[EA_{pr}(c')] \qquad \text{(Estimation by Prior)}$$
$$c = \arg\max_{c'\in\mathbb{Cr}}[EA_{c}(c')] \qquad \text{(Estimation by Posterior)}$$
$$c = \arg\max_{c'\in\mathbb{Cr}}[EA_{\pi^?}(c')], \pi^? \in \mathbb{Cr} \qquad \text{(Estimation by Something Else)}$$

Three approaches to $\mathbb{E}$:[5]

$$EA^{\perp\!\!\!\perp}_\pi(\phi[c']) =_{\text{df}} EA_\pi(c') \qquad \text{(Independent EDT)}$$
$$EA^{|}_\pi(\phi[c']) =_{\text{df}} EA_{\pi(\,\cdot\,|\phi[c'])}(c') \qquad \text{(Evidential EDT)}$$
$$EA^{\|}_\pi(\phi[c']) =_{\text{df}} EA_{\pi(\,\cdot\,\|\phi[c'])}(c') \qquad \text{(Causal EDT)}$$

[5] These are akin to standard (Savage (1954)), causal (Jeffrey (1996)), and evidential (Lewis (1981)) decision theory. Independent EDT weighs accuracy scores at w by c's credence that w is actual, Evidental EDT weighs them by c's credence in the subjunctive conditional probability that w is actual given that c' was adopted. Causal EDT weights them by c's credence in the indicative conditional probability that w would be actual were c' to be adopted.

The "default" is probably Estimation by Prior and Causal EDT.[6] The authors we're considering today opt for Independent EDT and Estimation by Something Else.

[6] Causal vs Evidential won't be relevant today, though it is important in other cases.

2 | Imps.

The Bribe: Hilary Greaves (2013) raises the following objection to EDT:

Imps: Emily is taking a walk through the Garden of Epistemic Imps. A child plays on the grass in front of her. In a nearby summerhouse are n further children, each of whom may or may not come out to play in a minute. They are able to read Emily's mind, and their algorithm for deciding whether to play outdoors is as follows. If she forms degree of belief 0 that there is now a child before her, they will come out to play. If she forms degree of belief 1 that there is a child before her, they will roll a fair die, and come out to play iff the outcome is an even number.

Let IMP_i for $i \in [1, n]$ be the proposition that the ith child will come out to play and let IMP_{now} be the proposition that there is a child in front of Emily. Then the specifications of the case place the following constraints on Emily's prior (pr):

$$pr(\text{IMP}_{now}) = 1, \quad pr(\text{IMP}_i \mid \langle \text{Emily adopts credence } x \text{ in } \text{IMP}_{now} \rangle) = 1 - x/2.$$

Emily can adopt any credence $x \in [0, 1]$ in IMP_{now}. For $x \in [0, 1)$, this amounts to taking the bribe to some degree. The exact degree to which EDT recommends that Emily drop her credence in IMP_{now} will depend on how we fill out the EDT Schema.[7]

Problem: in *Imps*, the Accuracy Norm conflicts with the **Evidentialist Norm** that when presented with compelling evidence that P is true, have high credence in P.

3 | K&L and Joyce Reconcile Accuracy and Evidentialism

Konek and Levinstein (2019) and Joyce (2018) recover the evidentialist intuition in *Imps* by using Independent EDT and changing the estimation credence function. On their theories, an agent with credence function c and evidence E is rational only if:[8]

1. **K&L's EDT**: $c = \arg\max_{c' \in \mathbb{C}\mathbb{r}} [EA^{\perp\!\!\!\perp}_{ch(\,\cdot\,|E \wedge \phi[c])}(c')]$.
2. **Joyce's EDT**: $c = \arg\max_{c' \in \mathbb{C}\mathbb{r}} [EA^{\perp\!\!\!\perp}_{pr(\,\cdot\,\|\phi[c])}(c')]$.

To derive these results, use *pairs* $\langle c, s \rangle$ notation of credence functions and world states:

> [EST→PREF]: For any pairs $\langle c, s \rangle$, $\langle c', s' \rangle$, scoring rule A, estimation method $\mathbb{E}$, and credence function π, say that $\langle c, s \rangle \trianglerighteq \langle c', s' \rangle$ (relative to A, $\mathbb{E}$, π) if $\langle c, s \rangle$ has greater expected accuracy: i.e. $EA^{\mathbb{E}}_{\pi(\,\cdot\,|s)}(c) \geq EA^{\mathbb{E}}_{\pi(\,\cdot\,|s')}(c')$.

In both theories, a credal state is *rational* only if it is *ratifiable* (i.e., optimal if chosen).[9]

> RatRat: It is epistemically rational for an agent S to choose to adopt credence function c only if c is **ratifiable**: for every alternative state b, $\langle c, s_c \rangle \trianglerighteq \langle b, s_c \rangle$, where s_c is the state that would obtain if S adopted c.

[EST→PREF] + K&L's or Joyce's choice of π are implementations of RatRat and deliver their versions of EDT and the evidentialist intuition in *Imps*.

Let em_x be the credence function with credence x in IMP_{now} and credence $1 - x/2$ in the IMP_i, and s_x be the state that obtains when Emily adopts em_x:

1. If Emily refuses the bribe $\phi[em_1]$, then she is rational because she prefers em_1 over all alternatives: $\langle em_1, s_1 \rangle \geq \langle b, s_1 \rangle$ for all $b \in \mathbb{C}\mathbb{r}$.
2. If Emily takes the bribe to any degree and $\phi[em_x]$ with $x \neq 1$, then she is irrational because she prefers the chance function: $\langle em_0, s_0 \rangle \triangleleft \langle ch(\,\cdot \mid s_0), s_0 \rangle$.

Basically, according to these EDTs, Emily is only rational if she matches the chance function, and she matches the chance function only if she refuses the bribe.

[7] If we only evaluate the accuracy of the propositions IMP_i, Greaves (2013) shows that EDT will recommend that Emily adopt credence 0 in IMP_{now}. Joyce and Weatherson (2019) show that if we look at every proposition in the algebra, Emily should take a more modest bribe.

[8] for K&L, this is an "iff." But it's problematic enough as a necessary condition. For Joyce, there is an additional necessary condition, which we will also ignore.

[9] Jeffrey (1996)

4 | Objections.

4.1 | *Preference over Impossible Pairs of Credences and World States*

The following is a plausible constraint on preference over pairs:[10]

> RESTRICTIVE ADMISSIBILITY: For credence functions p and p' and world states s and s', pairs $\langle p, s\rangle, \langle p', s'\rangle$ can stand in relation $\langle p, s\rangle \trianglerighteq \langle p', s'\rangle$ only if $\langle p, s\rangle$ and $\langle p', s'\rangle$ are both compatible with the setup of the case.

K&L and Joyce must deny RESTRICTIVE ADMISSIBILITY, because RATRAT requires Emily to have preferences over pairs $\langle p, s\rangle$ that are not compatible with the setup of the case.

Why? For any em_x, K&L and Joyce say Emily should prefer to adopt $ch(\cdot \mid s_x)$.[11]

→ Ex: Emily takes the bribe and adopts em_0. Then she should prefer $\langle ch(\cdot \mid s_0), s_0\rangle$ to $\langle em_0, s_0\rangle$. But $\langle ch(\cdot \mid s_0), s_0\rangle$ is incompatible with the setup of *Imps*.

So, if RESTRICTIVE ADMISSIBILITY holds, neither theory delivers the desired results.

[10] Carr (2017): "What is the value of adopting a credence at an outcome where no one adopts the credence function?" Pettigrew (2018): "[W]hether [an option] is rationally permissible [...] does not depend upon [the utilities of] options at worlds at which those options could not possibly be adopted."

[11] s_x depends only on $em(\text{IMP}_{now}) = x$, so s_x can include more finely grained possible worlds in which Emily has different credences in each IMP_i.

4.2 | *Is Ratifiability Properly Motivated?*

We argue that neither K&L nor Joyce give successful defenses of RATRAT.

K&L: K&L distinguish between epistemic *acts* and epistemic *states*.

1. Valuable epistemic **acts** *change* the world to be most accuracy-conducive.
2. Valuable epistemic **states** *reflect* the world the most accurately.

Claim: epistemic rationality is about *states* and not acts. Emily should prefer the *act* of taking the bribe, but not the resultant epistemic *state* (because it is unratifiable).

Concern: K&L say that an agent should prefer the c that *changes the world* so that there are no pairs $\langle c', s_c\rangle$ that the agent would prefer to adopt while holding fixed s_c.[12]

→ This is act-based: it says to influence the world to increase epistemic value.

[12] This derives from denying RESTRICTIVE ADMISSIBILITY. If impossible pairs are relevant to epistemic preference, but they cannot actually be chosen (since impossible), then it is rational to choose the option that eliminates the impossible pairs that are more accurate than one's own pair.

Joyce: Joyce argues that ratifiability identifies distinctively epistemic evaluation:

> I see [ratifiability] as distinguishing genuine epistemic choices, those in which the agent sees herself as choosing real credences, from sham-epistemic choices, in which she sees herself choosing sham "credences." The hallmark of real credences is that the believer is happy to use them as the basis for making estimates of truth-values and quantities that depend on truth-values, the characteristic things credences are meant to do. In contrast, believers with sham "credences" will not want to use them to do these jobs, but will instead aim to switch to alternative credences before making estimates of truth-values, evaluating bets, etc. (2018, p. 258).

The "Dutch Book": If agents adopt unratifiable states, they will place bets that are sure losses from the standpoint of the available evidence.[13]

Concern: Joyce says Emily shouldn't value "sham" credences for the purpose of estimating truth values. But she should value those credences–they make *better* overall estimates! She just can't value her credences in *every proposition* the right way.

[13] This is not a standard synchronic or diachronic Dutch Book argument. Here, Emily loses with certainty because her credences do not match her evidence, and the bet is settled by the evidence.

Concern: Reinterprets epistemic value as "improvement-minimization" rather than "accuracy-maximization": change the world so *there is no way to be more accurate* in the resultant state, rather than changing it so *you are most accurate* in the resultant state.[14]

→ Ratifiability is compelling. It looks bad to knowingly select an option is dominated. But in this case, the undominated options are less accurate overall.

An analogy: Choose the largest possible number from either {1, 2, 3} or {3, 4, 5}. We can choose freely from {1, 2, 3}, but we can *at most* select 4 from {3, 4, 5}. Choosing 3 from {1, 2, 3} gives us the largest number possible holding fixed the context. Choosing from {3, 4, 5} gives us the largest overal;, but is dominated in the context. Ratifiability tells us to choose 3 from the first set, but we should presumably choose 4 from the second.

[14] This is equally a problem for K&L: not only do they recommend changing the world after all, but it's not clear that they have the right recommendation for *how* to change it.

4.3 | *Bonus Objection: Bringing Chance to the Accuracy-First Party*

Is K&L's chance-deference assumption appropriate?[15] Two worries:

1. No clear accuracy-theoretic justification for using the chance function.
2. Without justification, it's an "accuracy + chance" rather than "accuracy" theory.

[15] NB: Joyce obtains a similar theory without needing the chance function. But even if K&L have a justification for using chance to estimate, they still require ratifiability.

5 | An Error Theory.

Consider two consequentialist goods that are at odds with each other in *Imps*:

- Maximize Local Accuracy: adopt a credence function c such that no other (relevant) credence function c' is more accurate on any proposition P.
- **Maximize Global Accuracy:** adopt a credence function c such that no other (relevant) credence function c' has greater global accuracy.

What EDT recommends to Emily depends on which good receives primacy. If we privilege global accuracy, then Emily ought to take the bribe. If we privilege local accuracy by adding ratifiability to the framework, then Emily ought to refuse.

Potential reasons to privilege local accuracy:

- **Value**: We might care more about being closer to the truth on certain propositions. E.g., a friend's birthday vs. the distance between NY and LA.
  Response: Ratifiability isn't the best way to build this into the framework.
- **Ought→Can**: Emily simply can't believe against the evidence.
  Response: We want to screen off concerns about voluntarism for our purposes.

Assume that Emily values each proposition equally and we screen off voluntarist worries. Then we think it's correct to say that Emily should take the bribe.

Notice: there are a number of everyday cases in which we knowingly (and rationally) sacrifice accuracy in one set of propositions to acquire greater accuracy elsewhere.

→ **Example:** We watch Seasons 1-2 of *Severance* to become more accurate about whether the internet conspiracies are correct, at the cost of having accurate credences about the events of the new season of *White Lotus*.[16]

This is a feature of having bounded epistemic capacities. The only difference between this case and Emily's is that we know *that* we have sacrificed accuracy; we just don't know *what* in particular is the nature of the sacrifice. But knowing *what* she sacrifices shouldn't make a difference to Emily.

[16] NB: neither of the authors has watched the new seasons of *Severance* or *White Lotus*.