Accuracy-Centered Epistemic Utility Theory

Julien Dutant[1] Branden Fitelson[2]

[1]Philosophy @ KCL

julien.dutant@kcl.ac.uk

[2]Philosophy & Religion @ Northeastern

branden@fitelson.org

Outline

- **Miners** [34, 26]. You are standing in front of two mine shafts (*A* and *B*). Flood waters are approaching. You know that ten miners are in one of the shafts, but you don't know which (*e.g.*, their location was determined by the result of a fair coin toss). You have enough sand bags to block one of the shafts. If the miners are in *A*, then blocking *A* saves all 10 miners (and, hence, minimizes disutility, *i.e.*, # of dead miners). If the miners are in *B*, then blocking *B* minimizes disutility. If you block neither *A* nor *B*, the water will be divided, and only the lowest miner in the shaft will die. **Claim**. *It is rationally permissible to block neither A nor B*.
- **Gibbard's Coin** [14, 30]. A fair coin has been tossed (and you have no information about how it landed). If it landed Heads (*H*), then believing *H* is the attitude which minimizes (epistemic) disutility (*viz.*, *inaccuracy*). If it landed Tails (*T*), then believing *T* is the attitude which minimizes inaccuracy. **Claim**. *It is rationally permissible to believe neither H nor T*.

☞ *It can be rationally permissible to (knowingly) occupy a state, which does **not** minimize disutility — in **any** possible world.*

- Today's talk is about (i) formal, (ii) synchronic, (iii) epistemic (iv) coherence (v) requirements (of ideal rationality).
  - (i) *Formal* coherence is to be distinguished from other sorts of coherence discussed in contemporary epistemology (*e.g.*, in some empirical, truth/knowledge-conducive sense [1]).
    - Our notions of coherence will supervene on *logical* (and *formal probabilistic*) properties of judgment sets.
  - (ii) *Synchronic* coherence has to do with the coherence of a set of judgments held by an agent *S at a single time t*.
    - So, we'll *not* be discussing any *diachronic* [40] requirements.
  - (iii) *Epistemic* coherence involves *distinctively* epistemic values (specifically: *accuracy* [19] and *evidential support* [7]).
    - This is to be distinguished from *pragmatic* coherence (*e.g.*, immunity from dutch books [38], and the like [17]).
  - (iv) *Coherence* has to do with how a set of judgments "hangs together". CRs are *wide-scope* [3], global requirements.
  - (v) *Requirements* are *evaluative*; they give *necessary* conditions for (ideal) epistemic rationality of a doxastic state [40].

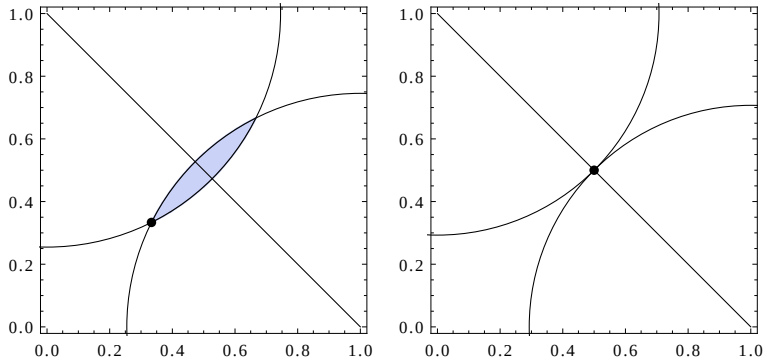
- Joyce [22, 20] gives epistemic utility arguments for *probabilism* as a coherence requirement for sets of numerical confidence judgments (*viz.*, credences).
- These arguments trace back to de Finetti [8] and they have recently culminated in powerful generalizations [37, 20].
- I will not go through these more recent arguments here (as I want to focus mainly on *belief*). See Richard Pettigrew's SEP entry [35] for a state-of-the-art reference on that dialectic.
- For my purposes today, it will suffice to explain how de Finetti's original argument works. This will give us a template for applying similar techniques to full belief.
- Let us suppose the following *alethic ideal* for credences (analogous to The Truth Norm implicit in **Gibbard's Coin**).

> **Alethic Ideal** (for credence). An agent's credence $b(p)$ should (*alethically, ideally*) be *maximal* [*i.e.*, $b(p) = 1$] if p is *true* and *minimal* [*i.e.*, $b(p) = 0$] if p is *false*.

- The Alethic Ideal is fine *as an ideal*, but it does *not* constitute a (general) *rational requirement*. This is (at least in part) because it tends to conflict with the Evidential Ideal.
  - **Evidential Ideal** (for credence). An agent's credence $b(p)$ should (*evidentially, ideally*) reflect the strength of an agent's total evidence regarding p.
- Consider the credal analogue of **Gibbard's Coin**. According to the Alethic ideal, if H is true, then $b(H)$ should equal 1; and, if T is true, then $b(T)$ should equal 1. But, the credal assignment $b(H) = b(T) = 1/2$ is rationally permissible.
- If you reflect on the sorts of cases we've been discussing so far, it is possible to garner a key insight about the nature of rational requirements (as we will understand them here).

☞ Rationality does not require the *actual* minimization of disutility (or even its *possible* minimization). It requires (something like) minimization of *expected* disutility.

- Decision-theoretic notions like expected disutility minimization (or non-dominance in utility, *etc.*) are essential to EUT arguments for rational (coherence) requirements.
- Let's consider possible assignments of credence to a contingent pair $\{P, \neg P\}$. We'll represent credal assignments $b(\cdot) \in [0,1]$ on $\{P, \neg P\}$, as: $b(P) = x$ and $b(\neg P) = y$.
- This allows us to visualize the salient space of possible credence functions on $\{P, \neg P\}$ *via* a simple Cartesian plot (of the unit square), with abscissa $b(P)$ and ordinate $b(\neg P)$.
- The *gradational inaccuracy* of $b(p)$ at a world w is a function $\mathfrak{i}_b$ of $b(p)$ and w's *indicator function* $v_w(p)$, which assigns 1 (0) to truths (falsehoods) in w. The *overall* inaccuracy $\mathfrak{I}_b$ of $b(\cdot)$ at w is $\mathfrak{I}_b(b, w) \stackrel{\text{def}}{=} \sum_p \mathfrak{i}_b(b(p), v_w(p))$.
- de Finetti's measure $\mathfrak{i}_b(b(p), v_w(p)) \stackrel{\text{def}}{=} (v_w(p) - b(p))^2$ is known as the Brier Score. This implies $\mathfrak{I}_b$ is the (squared) *Euclidean distance* between the vectors $b(\cdot)$ and $v_w(\cdot)$.



- **Simplest case of dF's Theorem [8]**. The diagonal lines are the *probabilistic* b's (on $\langle P, \neg P\rangle$). The point $\langle 1, 0\rangle$ ($\langle 0, 1\rangle$) corresponds to the world in which P is true (false).

> **Theorem** (de Finetti [8]). b is *non*-probabilistic $\Leftrightarrow \exists\, b'(\cdot)$ which is (Euclidean) *closer to* $v_w(\cdot)$ *in every possible world*.

- The plot on the left (right) explains the $\Rightarrow$ ($\Leftarrow$) direction.

- Here is a — perhaps *the* — "paradigm" CR [36, 39, 32, 23].
  - **The Consistency Requirement for Belief**. Agents should have *sets* of beliefs that are *logically consistent*.
- The Consistency Requirement is implied by The Alethic Ideal (*i.e.*, if S is Alethically Ideal, then S's beliefs are consistent).
  - **Alethic Ideal** (for belief). S should (*alethically, ideally*) believe (disbelieve) that p just in case p is true (false).
- We've already seen (**Gibbard's Coin**) that The Alethic Ideal can come into *conflict* with The Evidential Ideal.
  - **The Evidential Ideal** (for belief). S should (*evidentially, ideally*) believe (disbelieve) p if S's total evidence supports (counter-supports) p. Otherwise, S should *suspend* on p.
- More subtle cases reveal that The Consistency Requirement can also conflict with The Evidential Ideal [6, 25, 13, 24].
- We'll refer to the claim that there exist *some* such cases as *the datum*. Foley's [13] explanation of *the datum* is helpful.

"...if the avoidance of recognizable inconsistency were an absolute prerequisite of rational belief, we could not rationally believe each member of a set of propositions and also rationally believe of this set that at least one of its members is false. But this in turn pressures us to be unduly cautious. It pressures us to believe only those propositions that are certain or at least close to certain for us, since otherwise we are likely to have reasons to believe that at least one of these propositions is false. At first glance, the requirement that we avoid recognizable inconsistency seems little enough to ask in the name of rationality. It asks only that we avoid certain error. It turns out, however, that this is far too much to ask."

- We will offer a precise explication of Foley's position. The basic idea will be to work out the theoretical consequences of the slogan: *Epistemic rationality requires minimization of expected inaccuracy*. But, first, a First-Order Preface Case.

First-Order Preface Paradox ([10]). John is an excellent empirical scientist. He has devoted his entire (long and esteemed) scientific career to gathering and assessing the evidence that is relevant to the following first-order, empirical hypothesis: (H) all scientific/empirical books of sufficient complexity contain at least one false claim. By the end of his career, John is ready to publish his masterpiece, which is an exhaustive, encyclopedic, 15-volume (scientific/empirical) book which aims to summarize (all) the evidence that contemporary empirical science takes to be relevant to H. John sits down to write the Preface to his masterpiece. Rather than reflecting on his own fallibility, John simply reflects on the contents of (the main text of) his book, which constitutes *very strong inductive evidence in favor of* H. On this basis, John (inductively) infers H. But, John also believes each of the individual claims asserted in the main text of the book. Thus, because John believes (indeed, knows) that his masterpiece instantiates the antecedent of H, the (total) set of John's (rational/justified) beliefs is inconsistent.

- We assume that our agent has a credence function $b(\cdot)$, which is *probabilistic*. This allows us to use $b(\cdot)$ to define notions of (subjective) *expected* (epistemic) utility.
- We assume that our agent takes exactly one of three qualitative attitudes (B, D, S) toward each member of a finite agenda $\mathcal{A}$ of (classical, possible worlds) propositions.
- We do *not* assume that these qualitative judgments can be *reduced* to $b(\cdot)$. But, we will use $b(\cdot)$ to derive a *rational coherence constraint* for qualitative judgment sets $\mathbf{B}$ (on $\mathcal{A}$).
- This derivation requires both the agent's credence function $b(\cdot)$ and their *epistemic utility function* [18, 29, 31] $u(\cdot)$.
  - ☞ Following Easwaran [11] & Dorst [9], we assume our agent cares *only* about whether their judgments are *accurate*.
- Specifically, our agent attaches some *positive* utility ($\mathfrak{r}$) with making an *accurate* judgment, and some *negative* utility ($-\mathfrak{w}$) with making an *inaccurate* judgment (where $\mathfrak{w} > \mathfrak{r} > 0$).

- Because suspensions are neither accurate nor inaccurate (*per se*), our agent will attach *zero* epistemic utility to suspensions $S(p)$, independently of the truth-value of p.
- Thus, we have the following piecewise definition of $u(\cdot, w)$.

$$u(B(p), w) \stackrel{\text{def}}{=} \begin{cases} -\mathfrak{w} & \text{if } p \text{ is false at } w \\ \mathfrak{r} & \text{if } p \text{ is true at } w \end{cases}$$

$$u(D(p), w) \stackrel{\text{def}}{=} \begin{cases} \mathfrak{r} & \text{if } p \text{ is false at } w \\ -\mathfrak{w} & \text{if } p \text{ is true at } w \end{cases}$$

$$u(S(p), w) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } p \text{ is false at } w \\ 0 & \text{if } p \text{ is true at } w \end{cases}$$

- With this *accuracy-centered* epistemic utility function in hand, we can derive a naïve EUT coherence requirement.

- To do so, we'll also need a *decision-theoretic principle*.
- As we saw, applications of EUT to grounding probabilism as a (synchronic) requirement for $b(\cdot)$ typically appeal to a *non-dominance* (in epistemic utility) principle [20, 37, 35].
- But, some authors apply an *expected epistemic utility maximization* (or *expected inaccuracy minimization*) principle to derive rational requirements [28, 16, 12, 33].

> **Coherence**. An agent's belief set $\mathbf{B}$ over an agenda $\mathcal{A}$ should, from the point of view of their own credence function $b(\cdot)$, *maximize expected epistemic utility* (or *minimize expected inaccuracy*). That is, $\mathbf{B}$ should maximize
>
> $$EEU(\mathbf{B}, b) \stackrel{\text{def}}{=} \sum_{p \in \mathcal{A}} \sum_{w \in W} b(w) \cdot u(\mathbf{B}(p), w)$$
>
> where $\mathbf{B}(p)$ is the agent's attitude toward p, and $W \stackrel{\text{def}}{=} \bigcup \mathcal{A}$.

- We also assume "*act-state independence*": $\mathbf{B}(p)$ and p are *b-independent* [15, 5, 4, 27]. See Extras for discussion.

- The consequences of **Coherence** are rather simple and intuitive. It is straightforward to prove the following result.

> **Theorem** ([11, 9]). An agent with credence function $b(\cdot)$ and qualitative judgment set $\mathbf{B}$ over agenda $\mathcal{A}$ satisfies **Coherence** *if and only if* for all $p \in \mathcal{A}$
>
> $$B(p) \in \mathbf{B} \textit{ iff } b(p) > \tfrac{\mathfrak{w}}{\mathfrak{r}+\mathfrak{w}},$$
> $$D(p) \in \mathbf{B} \textit{ iff } b(p) < \tfrac{\mathfrak{r}}{\mathfrak{r}+\mathfrak{w}},$$
> $$S(p) \in \mathbf{B} \textit{ iff } b(p) \in \left[\tfrac{\mathfrak{r}}{\mathfrak{r}+\mathfrak{w}}, \tfrac{\mathfrak{w}}{\mathfrak{r}+\mathfrak{w}}\right].$$

- In other words, **Coherence** *entails Lockean representability*, where the Lockean thresholds are determined by the way the agent (relatively) values accuracy *vs.* inaccuracy.
- This provides an elegant, EUT-based explanation of why Lockean representability is a rational requirement for agents with *both* credences *and* qualitative attitudes.

☞ As Dorst [9] puts it: *Lockeans maximize expected accuracy*.

- By way of summary, it is useful to think about the analogy between the norms we've been discussing, and principles of rational choice theory: **The Decision-Theoretic Analogy**.

Epistemic Principle	Analogous Decision-Theoretic Principle
Alethic Ideal	(AMU) Do ϕ only if ϕ maximizes utility in the *actual* world.
Consistency	(PMU) Do ϕ only if ϕ maximizes u in *some possible* world.
Coherence	(MEU) Do ϕ only if ϕ maximizes EU (relative to *some* Pr).
(WADA)	(WDOM) Do ϕ only if ϕ is *not weakly dominated* in utility.
(SADA)	(SDOM) Do ϕ only if ϕ is *not strictly dominated* in utility.

- Like the **Alethic Ideal**, (AMU) is *not* a *requirement of rationality*; and, like **Consistency**, (PMU) isn't a rational requirement either (this was the lesson of **Miners** [34, 26]).
- As Foley (*op. cit.*) explains, **Consistency** is *too demanding*. But, **Coherence** is *not* — it does *not* "pressure us to believe only those propositions that are (close to) certain for us".

- Suppose our (naïve) agent has a belief set $\mathbf{B}_n$ on a *minimal inconsistent* agenda of size n (*e.g.*, $(n-1)$–ticket lottery).

  Theorem ([10]). For all $n \geq 2$ and any probability function $\Pr(\cdot)$, the $\Pr(\cdot)$-Lockean-representability of $\mathbf{B}_n$ (with threshold t) *entails* deductive consistency of $\mathbf{B}_n$ *iff* $t \geq \frac{n-1}{n}$.

- If we combine this with Easwaran's **Coherence** theorem, we get the following result, regarding the conditions under which the **Coherence** of $\mathbf{B}_n$ *entails* the *consistency* of $\mathbf{B}_n$.

  Theorem. For all $n \geq 2$, an agent with an accuracy-centered utility function u, a credence function $b(\cdot)$, and a belief set $\mathbf{B}_n$, the **Coherence** of $\mathbf{B}_n$ entails the consistency of $\mathbf{B}_n$ *iff*

  (†) $$\mathfrak{w} \geq (n-1) \cdot \mathfrak{r}.$$

☞ Insisting that **Coherence** *implies consistency* (wrt $\mathbf{B}_n$) *requires (naïve) agents to disvalue inaccuracy at least* $(n-1)$ *times as much as they value accuracy.*

- We've been presenting epistemic requirements as if they applied to "doxastic acts" of believing, disbelieving or suspending judgment (or assigning some credence).
- Strictly speaking, we should present both epistemic and prudential requirements as constraints on *preferences*.
- For instance, the key evaluative claim about **Miners** is (strictly speaking) that the (partial) *preference ranking*

  $$\mathsf{C} \succ \mathsf{A} \sim \mathsf{B}$$

  is not irrational — because it is aligned with the agent's expected utility ranking (where $\mathsf{C} \triangleq$ blocking neither shaft).
- Similarly, the key evaluative claim about **Gibbard's Coin** is (strictly speaking) that the (partial) *preference ranking*

  $$\mathsf{S} \succ \mathsf{H} \sim \mathsf{T}$$

  is not irrational — since it is aligned with expected *epistemic* utility (where $\mathsf{S} \triangleq$ believing neither H nor T).

- If an agent does not have (precise) credences, expected inaccuracy minimization will not be an apt coherence requirement. But, we can still say *something* here.
- We can appeal to *non-dominance* requirements, such as:

  Weak Accuracy-Dominance Avoidance (WADA).

  There does *not* exist an alternative belief set $\mathbf{B}'$ such that:
  (i) $(\forall w)[u(\mathbf{B}', w) \leq u(\mathbf{B}, w)]$, and
  (ii) $(\exists w)[u(\mathbf{B}', w) < u(\mathbf{B}, w)]$.

  Strict Accuracy-Dominance Avoidance (SADA).

  There does *not* exist an alternative belief set $\mathbf{B}'$ such that:
  (iii) $(\forall w)[u(\mathbf{B}', w) < u(\mathbf{B}, w)]$.

- It turns out [10, 11] that **Coherence** ⇒ (WADA) ⇒ (SADA).
- Indeed, (WADA) and (SADA) are *very* weak [10]. But, they do constitute non-trivial *necessary requirements* of rationality.

- Sharon Ryan [39] gives an argument *for* (CB) as a rational requirement, which makes use of these three premises.

  The Closure of Rational Belief Principle (CRBP).
  If S rationally believes p at t and S knows (at t) that p entails q, then it would be rational for S to believe q at t.

  The No Known Contradictions Principle (NKCP).
  If S knows (at t) that $\bot$ is a logical contradiction, then it would *not* be rational for S to believe $\bot$ (at t).

  The Conjunction Principle (CP).
  If S rationally believes p at t and S rationally believes q at t, then it would be rational for S to believe $\ulcorner p \,\&\, q \urcorner$ at t.

- Ryan's (CRBP) & (NKCP) have analogues in our framework (which *are* coherence requirements). But, (CP) does *not*.

(SPC) If $p \vDash q$, then any $\mathbf{B}$ s.t. $\{B(p), D(q)\} \subseteq \mathbf{B}$ is incoherent.
(NCB) Any $\mathbf{B}$ such that $\{B(\bot)\} \subseteq \mathbf{B}$ is incoherent.
¬(CP) *Not* every $\mathbf{B}$ s.t. $\{B(p), B(q), D(p \,\&\, q)\} \subseteq \mathbf{B}$ is incoherent.

- We mentioned above that we assume "act-state independence" (ASI). There are two main reasons for (ASI).
- If $\mathbf{B}(p)$ and p are correlated under $b(\cdot)$, then the verdicts delivered by **Coherence** can be *partition-sensitive*, *i.e.*, they can depend on the way in which the underlying set of doxastic possibilities is partitioned or carved up [21].
- More importantly, if $\mathbf{B}(p)$ and p are correlated under $b(\cdot)$, then EUT can yield unintuitive (and/or odd) verdicts (even assuming a "natural" partition of states). See [4, 15, 5, 27].
- For instance, Carr [5] considers cases in which $B(p)$ and p are *positively* correlated (*e.g.*, believing you will do a handstand makes it much more likely that you will).
- Examples involving *negative* correlation between $B(p)$ and p have been discussed by various authors (*e.g.*, [15]). The most extreme (and difficult) examples along these lines are the self-referential examples due to Michael Caie [4].

- Caie's original example involved (only) *credences* [4]. It was designed to undermine Joycean (accuracy-dominance) arguments for *probabilism* as a requirement for $b(\cdot)$.
- There are analogous examples for full belief. Consider:

  (P) S does not believe that P. $[\neg B(\ulcorner P \urcorner).]$

- One can argue (Caie-style) that the only non-dominated (opinionated) belief sets on $\{P, \neg P\}$ are $\{B(P), B(\neg P)\}$ and $\{D(P), D(\neg P)\}$, which are both *ruled-out* by **Coherence**.

	P	$\neg P$	$B(P)$	$B(\neg P)$	$B(P)$	$D(\neg P)$	$D(P)$	$B(\neg P)$	$D(P)$	$D(\neg P)$
w_1	**F**	**T**	$-$	$+$	$-$	$-$	$\times$	$\times$	$\times$	$\times$
w_2	**T**	**F**	$\times$	$\times$	$\times$	$\times$	$-$	$-$	$-$	$+$

- The "$\times$"s indicate that these worlds are *ruled-out* (*a priori*) by the definition of P. As such, the only non-dominated belief sets seem to be $\{B(P), B(\neg P)\}$ and $\{D(P), D(\neg P)\}$.
- If this Caie-style reasoning is correct, then it shows that *some of our assumptions must go*. But, which one(s)?

[1] L. Bonjour, *The Coherence Theory of Empirical Knowledge*, *Phil. Studies*, 1975.

[2] R. Briggs, F. Cariani, K. Easwaran, B. Fitelson, *Individual Coherence and Group Coherence*, *Essays in Collective Epistemology*, J. Lackey (*ed.*), OUP, 2014.

[3] J. Broome, *Wide or Narrow Scope*, *Mind*, 2007.

[4] M. Caie, *Rational Probabilistic Incoherence*, *Philosophical Review*, 2013.

[5] J. Carr, *How to Expect When You're Expecting*, 2014.

[6] D. Christensen, *Putting Logic in its Place*, OUP, 2007.

[7] W. Clifford, *The ethics of belief*, *Contemporary Review*, 1877.

[8] B. de Finetti, *The Theory of Probability*, Wiley, 1974.

[9] K. Dorst, *Lockeans Maximize Expected Accuracy*, 2015.

[10] K. Easwaran and B. Fitelson, *Accuracy, Coherence, and Evidence*, in *Oxford Studies in Epistemology V*, T. Szabo-Gendler and J. Hawthorne (*eds.*), OUP, 2015.

[11] K. Easwaran, *Dr. Truthlove or: How I Learned to Stop Worrying and Love Bayesian Probability*, *Noûs*, 2015.

[12] _____, *Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection*, *Philosophy of Science*, 2013.

[13] R. Foley, *Working Without a Net*, Oxford University Press, 1992.

[14] A. Gibbard, *Truth and correct belief*, *Philosophical Issues*, 2005.

[15] H. Greaves, *Epistemic Decision Theory*, *Mind*, 2013.

[16] H. Greaves and D. Wallace, *Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility*, *Mind*, 2006.

[17] A. Hájek, *Arguments for — or Against — Probabilism?*, *BJPS*, 2008.

[18] C. Hempel, *Deductive-Nomological vs. Statistical Explanation*, 1962.

[19] W. James, *The will to believe*, *The New World*, 1896.

[20] J. Joyce, *Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief*, in F. Huber and C. Schmidt-Petri (*eds.*), *Degrees of Belief*, 2009.

[21] _____, *The Foundations of Causal Decision Theory*, CUP, 1999.

[22] _____, *A Nonpragmatic Vindication of Probabilism*, *Philosophy of Science*, 1998.

[23] M. Kaplan, *Coming to Terms with our Human Fallibility*, *PPR*, 2013.

[24] P. Klein, *The virtues of inconsistency*, *The Monist*, 1985.

[25] N. Kolodny, *How Does Coherence Matter?*, *Proc. of the Aristotelian Society*, 2007.

[26] N. Kolodny and J. MacFarlane, *Ifs and oughts*, *Journal of Philosophy*, 2010.

[27] J. Konek and B. Levinstein, *The Foundations of Epistemic Decision Theory*, 2014.

[28] H. Leitgeb and R. Pettigrew, *An Objective Justification of Bayesianism I & II*, *Philosophy of Science*, 2010.

[29] I. Levi, *Gambling with Truth*, 1967.

[30] J. MacFarlane, *Assessment Sensitivity: Relative Truth and its Applications*, OUP, 2014.

[31] P. Maher, *Betting on Theories*, 1993.

[32] D. Nelkin, *The lottery paradox, knowledge, and rationality*, *Phil. Review*, 2000.

[33] G. Oddie, *Conditionalization, cogency, and cognitive value*, *BJPS*, 1997.

[34] D. Parfit, *What we together do*, unpublished manuscript, 1988.

[35] _____, *Epistemic Utility Arguments for Probabilism*, *SEP*, 2011.

[36] J. Pollock, *Nomic probability and the foundations of induction*, OUP, 1990.

[37] J. Predd, R. Seringer, E. Loeb, D. Osherson, H.V. Poor and S., Kulkarni, *Probabilistic coherence and proper scoring rules*, *IEEE Transactions*, 2009.

[38] F. Ramsey, *Truth and Probability*, 1926.

[39] S. Ryan, *The epistemic virtues of consistency*, *Synthese*, 1996.

[40] M. Titelbaum, *Quitting Certainties*, OUP, 2013.

Knowledge-Centered Epistemic Utility Theory

Julien Dutant[1] Branden Fitelson[2]

[1]Philosophy @ KCL

julien.dutant@kcl.ac.uk

[2]Philosophy & Religion @ Northeastern

branden@fitelson.org

Outline

- Epistemic utility theory (EUT) derives norms from *rational requirements* combined with a *theory of epistemic value*.
- So far EUT for beliefs has used the *Veritistic* theory of value: beliefs are good iff accurate or true [3, 2].
- But there is an alternative way of valuing beliefs:

  > **Plato's *Meno*.** For true opinions, as long as they remain, are a fine thing and all they do is good, but they are not willing to remain long, and they escape from a man?s mind, so that they are not worth much until one ties them down . . . That is why knowledge is prized higher than correct opinion, and knowledge differs from correct opinion in being tied down

- Gnosticism [13]: knowing is good. That is, believing p is good if you (thereby) know p.
- What do epistemic norms for beliefs look like on a Gnostic theory of value?

- Preliminary hurdle for gnosticism: act-state dependence.
  - We assume that knowledge requires belief.
  - So the states of our decision problems (whether one knows) will depend on the acts (whether one believes).
  - Worse: if you're certain that knowledge requires belief and that you do not believe p, then you're certain that you don't know p and the expected gnostic value of believing p is zero — can't recommend beliefs you do not already have.
- Two solutions:
  (1) Deny that knowledge requires belief [22, 19, 12].
  (2) Frame states in terms of *being in a position to know* [21].
    - Whether you are in a position to know p does not depend on whether you believe p.
    - If you are in a position to know p, then if you (come to) believe p (in the right way) you (thereby come to) know p.
- We use the second. *It is good to believe p if you are in a position to know p.* (Refined version to be introduced later.)

- Let Kp abbreviate *one is in a position to know* p.
  - We assume factivity: $Kp \to p$, and nothing else.
- As in Veritism we assume that *not* believing has the same value no matter how the world is. We set it to 0.

world w	$u(Bp, w)$	$u(\neg Bp, w)$
Kp	$\mathfrak{r}$	0
$p \,\&\, \neg Kp$	$\mathfrak{t}$	0
$\neg p$	$-\mathfrak{w}$	0

- Obvious constraints: $\mathfrak{r} \geq \mathfrak{t} \geq -\mathfrak{w}$ and $-\mathfrak{w} < 0$.
- Natural constraint: $\mathfrak{w} > \mathfrak{r}$.

  If $\mathfrak{w} < \mathfrak{t}$ we never allow not believing either of p and $\neg p$.

  If $\mathfrak{w} < \mathfrak{r}$ we sometimes allow believing both p and $\neg p$.

  $\mathfrak{w} > \mathfrak{r}$ will give a natural threshold ($1/2$).
- **Choice point**. $\mathfrak{t}$ — the value of mere true belief.

world w	$u(Bp, w)$	$u(\neg Bp, w)$
Kp	$\mathfrak{r}$	0
$p \,\&\, \neg Kp$	$\mathfrak{t}$	0
$\neg p$	$-\mathfrak{w}$	0

- **Pure Veritism**. $\mathfrak{t} = \mathfrak{r}$.
  Only truth matters.
- **Veritist Gnosticism**. $0 < \mathfrak{t} < \mathfrak{r}$.
  True belief is better than no belief, but not as good as knowledge.
- **Impure Gnosticism**. $-\mathfrak{w} < \mathfrak{t} < 0$.
  True belief is worse than no belief, but not as bad as false belief.
- **Pure Gnosticism**. $\mathfrak{t} = -\mathfrak{w}$.
  Only knowledge matters.

- Same epistemic utility theory framework as before:
- Our agent has a credence function $b(\cdot)$, which is *probabilistic*.
- They form a belief set $\mathbf{B}$ over a finite agenda $\mathcal{A}$ of (classical, possible worlds) propositions.
- We derive our requirements using the *epistemic utility maximization* principle [11, 7, 5, 18].

  Expectabilist Rational Requirement (ERR). An agent's belief set $\mathbf{B}$ (over $\mathcal{A}$) should, from the point of view of their credence function $b(\cdot)$, *maximize expected epistemic utility*. That is, $\mathbf{B}$ should maximize

$$EEU(\mathbf{B}, b) \stackrel{\text{def}}{=} \sum_{p \in \mathbf{B}} \sum_{w \in W} b(w) \cdot u(Bp, w)$$

- Assumes $u(\neg Bp, w) = 0$ so we simply sum over beliefs.
- Assumes act-state independence.

- Focus on Pure Gnosticism first. It yields the following epistemic utility function $u(\cdot, w)$:

$$u(Bp, w) = \begin{cases} -\mathfrak{w} & \text{if } Kp \text{ is false at } w \\ \mathfrak{r} & \text{if } Kp \text{ is true at } w \end{cases}$$

$$u(\neg Bp, w) = 0 \quad \text{at any } w$$

- We get the following result (analogue to [3, 2]):

  New Lockeanism. An agent with credence function $b(\cdot)$ and belief set $\mathbf{B}$ (over $\mathcal{A}$) satisfies (ERR) iff for all $p \in \mathcal{A}$

$$p \in \mathbf{B} \quad \textit{iff} \quad b(Kp) \geq \tfrac{\mathfrak{w}}{\mathfrak{r}+\mathfrak{w}} > 1/2$$

- Old Lockeanism.

  Veritistic Rational Requirement: Maximize the *expected truth-worth* of your beliefs (as gauged by your credences)!

  Believe p iff it is sufficiently likely that p *is true*.

  For instance, if $\mathfrak{w} = 2\mathfrak{r}$, believe p iff $b(p) > 2/3$.

- New Lockeanism.

  (Pure) Gnostic Rational Requirement: Maximize the *expected knowledge-worth* of your beliefs (as gauged by your credences)!

  Believe p iff it is sufficiently likely that p *is something you are in a position to know*.

  For instance, if $\mathfrak{w} = 2\mathfrak{r}$, believe p iff $b(Kp) > 2/3$.

- Now let us compare the normative implications of three views:
  - Old Lockeanism.
  - New Lockeanism.
  - Deontic Knowledge Norm: believe p iff you are in a position to know p.
- Example: all three views prohibit having pairwise contradictory beliefs. Suppose p and q are logically incompatible:
  - Old Lockeanism. By probabilism, either $b(p) < 1/2$ or $b(q) < 1/2$.
  - New Lockeanism. By probabilism and factivity, either $b(Kp) < 1/2$ or $b(Kq) < 1/2$.
  - Deontic Knowledge norm: by factivity, either $\neg Kp$ or $\neg Kq$.

- **Lottery and Preface paradoxes**.
  - **Lottery** [8]. A fair lottery with n tickets has been drawn but the winner is not yet announced. Let $p_1, p_2, \ldots, p_n$ be the propositions that ticket $1, 2, \ldots n$ lost.
  - **Preface** [15, 4]. An author writes a long and carefully researched history of publishing arguing that every book they examined contained at least one error. In the light of their research, they write in the preface that their own book contains at least one error.
    Let p be the preface claim and $p_2, p_3, \ldots, p_n$ the rest of the book's claim.
- In both cases, each p_i is highly likely but their conjunction is certainly false.
- Old Lockeanism: believe each p_i, but not large conjunctions of them. (*Inconsistent belief set*).
- Deontic Knowledge Norm, New Lockeanism? it depends on what one (is likely to) know in the cases.

- What do we know in the lottery and preface? A common view is that they contrast:
  - **Lottery**. In the lottery *one is not in a position to know any* p_i [21]. Sensitivity, safety absent.
  - **Preface**. In the preface the author is in a position to know some p_i. Sensitivity, safety may be satisfied.
  - 'Defeatist' views hold that evidence for the preface claim defeats knowledge of any claim in the book. They are on a sceptical slippery slope, for we know that some of our beliefs are false.
- Suppose the contrast is right. The Deontic Knowledge Norm then says:
  - **Lottery**. Do not believe any of the p_i.
  - **Preface**. Believe just those p_i (and conjunctions) that you (are in a position to) know.

- Now suppose that our agents go along with the contrast:
  - In the lottery, $b(Kp_i) \approx 0$ for any p_i.
  - in the preface, $b(Kp_i)$ is high for any p_i.
- Under these assumptions the New Lockeanism says:
  - **Lottery**. Do not believe any of the p_i.
  - **Preface**. Believe each p_i, but not large conjunctions of them. (*Inconsistent belief set*).
- So if there is a contrast in knowledge between the lottery and the preface, New Lockeanism:
  - sides with the Deontic Knowledge Norm re lottery: no belief.
  - sides with Old Lockeanism re preface: belief, inconsistency.
  - opposes Old Lockeanism in treating the cases differently.
  - opposes the Deontic Knowledge Norm in tolerating inconsistency.

- **Moore's paradox**. One believes both p and *I don't know* p.
- Old Lockeanism. p and *I don't know* p may both be very likely, *e.g.* in a lottery case.

☞ Recommends believing both: "The ticket will lose. I don't know that it will."

- Deontic Knowledge Norm. By factivity, one cannot know both. So one should not believe both.
- New Lockeanism. So far does not rule out believing both:
  - Suppose p is something you are in a position to know but do not actually know because you failed to think it through. Arguably you are *also* in a position to know, by noticing your lack of belief, that *you don't know* p.
  - Moreover, you could be in a position to know p, believe p, but do so for the wrong reasons and still fail to know p.
  - So one could believe both p and *I don't know* p and be in a position to know both ($2\mathfrak{r} > 0$).

- The case shows that we should refine our Gnostic theory:

☞ One should not value *believing* p *while one is in a position to know* p when one does so *without thereby knowing* p.

- **Refined Gnostic theory**. Refine our actions to be *believing* p ***on basis b*** $B(b,p)$.
  - Make the bases mutually exclusive: any relevant difference in basis counts as a different basis.
- Categorize our states in terms of *being in a position to know* p ***on basis b*** $B(b,p)$.
- For any basis b, if one believes p on basis b and one is in a position to know p on basis b then one knows p.

☞ That is all we need to say about bases: they correlate beliefs and one's epistemic position in the way described by that conditional.

- **Refined Pure Gnostic utility**:

$$u(B(b,p),w) = \begin{cases} -\mathfrak{w} & \text{if } K(b,p) \text{ is false at } w \\ \mathfrak{r} & \text{if } K(b,p) \text{ is true at } w \end{cases}$$

- Hence for any p, b: if one believes p on basis b, either one is not in a position to know p on basis b, or it is not the case that one fails to know p.
- Either way, the value of believing both p and *I don't know* p is at most $\mathfrak{r} - \mathfrak{w} < 0$. So it is dominated by suspending on both.
- So Refined New Lockeanism prohibits Moore-paradoxical beliefs.

- **Statistical Evidence in Law** [9, 20, 6].
  - **Testimony**. A witness claims seeing John climb over the fence to attend the rodeo without paying. In similar test conditions the witness is fairly, but not perfectly, reliable at identifying John.
  - **Statistics**. We know from ticket sales that most of the rodeo audience did not pay. We have proof that John was in the audience.
- Civil cases: one must be proved liable "by a preponderance of the evidence" or "by a balance of the probabilities". Often glossed as *more likely to be liable than not*.
- In both Testimony and Statistics our credence that John gate-crashed should be above $1/2$.
- However jurisprudence (and intuition) says that John could be found liable in Testimony but not in Statistics.

- A similar puzzle arises in *criminal* law [17, 16].
  - **Prison Yard Statistics**. A prison guard has been attacked and killed in a yard where twenty-five identically-dressed prisoners were standing. The sole witness saw that one, and only one, of the prisoners did not take part in the attack. But they could not identify the prisoner's individual features.
  - **Prison Yard Testimony**. As above, except that the sole witness was able to distinguish individual features and claims they can identify the innocent prisoner.
- Criminal case: one must be proved liable "beyond reasonable doubt". Often granted that that does not require "absolute certainty".
- Here as well, jurisprudence (and intuition) says that one could condemn in Testimony but not Statistics.

- Laudan [10] endorses a Old Lockean diagnosis. Set the price of accurate judgement and the cost of error, adopt the resulting threshold. At equal probabilities we should treat Statistics and Testimony equally.
- Contrast Blome-Tillmann's diagnosis [1]:
  - *a balance of the probabilities* is meant to require that it is more likely than not *that one knows*.
  - *beyond reasonable doubt* is meant to require that it is highly likely *that one knows*.

  ☞ Neo-Lockean standard: sufficiently likely *that one knows*.
  - In Testimony one is likely to be in a position to know p. [If the testifier is reliable, one's belief is safe, sensitive, *etc.*]
  - In Statistics one is arguably not in a position to know that p.
- The solution explains (a) why Statistics is not enough in either case, (b) how the criminal standard can be more stringent.

- Neo-Lockeanism can justify Blome-Tillmann's diagnosis:
  - One must believe guilty only if it is sufficiently likely that one knows.
  - How likely is sufficiently likely depends on the relative disvalue of *condemning without knowing* ($-\mathfrak{w}$). That is greater in the criminal case.
- Why would it make sense for the judicial system to adopt that standard? Littlejohn [14]:
  - If it only aimed at maximizing *accurate convinction*, it should follow Old Lockeanism and accept some level of statistical evidence.
  - If it aims at *retributing* and retributing requires knowledge (one can retribute for p only if one knows that p), it should follow New Lockeanism and not rely of evidence that fails to deliver knowledge.

- What about hybrid views: Impure Gnosticism and Veritist Gnosticism?
- In both views we get a complex threshold: the more likely p is, the less likely it needs to be than one *knows* that p for belief to be rational.
- Impure Gnosticism (true belief worse than no belief) sides with Pure Gnosticism. No belief in Moore's paradox, lotteries, statistical gatecrasher, but belief in the Preface.
- Veritist Gnosticism (true belief better than no belief) sides with Veritism. For some credences it will recommend Moore pardoxical beliefs, lottery beliefs and statistical gatecrasher beliefs.

- Summary of our applications:

	Old Lockean	New Lockean	Knowledge Norm
contradictory pair	✗	✗	✗
Moore paradox	✓	✗	✗
Lottery beliefs	✓	✗	✗
Statistical gatecrasher	✓	✗	✗
Preface beliefs	✓	✓	✗

- Neo-Lockeanism falls between Old Lockean and the Deontic Knowledge Norm.
- We find the package of verdicts attractive.

[1] M. Bloome-Tillman, *Sensitivity, Causality, and Statistical Evidence in Courts of Law*, *Thought*, 2015.

[2] K. Dorst, *Lockeans Maximize Expected Accuracy*, 2015.

[3] K. Easwaran, *Dr. Truthlove or: How I Learned to Stop Worrying and Love Bayesian Probability*, *Noûs*, 2015.

[4] K. Easwaran and B. Fitelson, *Accuracy, Coherence, and Evidence*, in *Oxford Studies in Epistemology V*, T. Szabo-Gendler and J. Hawthorne (*eds.*), OUP, 2015.

[5] ———, *Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection*, *Philosophy of Science*, 2013.

[6] D. Enoch and T. Fisher. *Sense and ?Sensitivity?: Epistemic and Instrumental Approaches to Statistical Evidence*, *Stanford Law Review*, 2015.

[7] H. Greaves and D. Wallace, *Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility*, *Mind*, 2006.

[8] H. Kyburg, *Probability and the Logic of Rational Belief*, Wesleyan University Press, 1961.

[9] H.H. Lai, *A Philosophy of Evidence Law: Justice in the Search for Truth*. Oxford University Press, 2008.

[10] L. Laudan *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*. Cambridge University Press, 2006.

[11] H. Leitgeb and R. Pettigrew, *An Objective Justification of Bayesianism I & II*, *Philosophy of Science*, 2010.

[12] D.K. Lewis. *Elusive Knowledge*, *Australasian Journal of Philosophy*, 1996.

[13] C. Littlejohn. *The Truth in Gnosticism*, *Análisis, Revista de Investigación Filosófica*, forth.

[14] C. Littlejohn, *Truth, Knowledge, and the Standard of Proof in Criminal Law*, *Synthese*, forth.

[15] D. Makinson. *The Paradox of the Preface*, *Analysis*, 1965.

[16] S. Moss, *Probabilistic Knowledge*, Oxford University Press, 2018.

[17] C. Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, *Harvard Law Review*, 1979.

[18] G. Oddie, *Conditionalization, cogency, and cognitive value*, *BJPS*, 1997.

[19] C. Radford, *Knowledge — by examples*, *Analysis*, 1966.

[20] M. Redmayne, *Exploring the Proof Paradoxes*, *Legal Theory*, 2008.

[21] T. Williamson, *Knowledge and its Limits*, Oxford University Press, 2000.

[22] A.D. Woozley, *Knowledge and not Knowing*, *Proceedings of the Aristotelian Society*, 1953.