Lecture #1: Credence & Evidential Probability

Branden Fitelson[1]

Rutgers Philosophy
&
MCMP @ LMU

branden@fitelson.org
http://fitelson.org/

[1]These lectures draw on joint work with Rachael Briggs (ANU), Fabrizio Cariani (NU), Kenny Easwaran (USC), and David McCarthy (HKU).

- The seminar will meet three times this first week (1–3pm, T/W/R). I'll give 3 lectures—an overview of my book project.
- Next week, we will have (a) a Q&A session on Tuesday (1–3), and then (b) a "project pitch" session on Thursday (1–3).
  - The Q&A session (Tuesday 6/10) will be an opportunity for you to ask questions to help shape your projects.
  - Then, on Thursday 6/12, individuals (or project groups of 2) will give a 15-minute "pitch" of their project proposals.
- Finally, we will have (c) 45-minute presentations of student projects, and (d) 10-page project reports will be due on the last class day (6/26). We have two options regarding (c).

(c_1) Some presentations on 6/17–6/18 & some in the final week.

(c_2) Or, we could use 6/18 (Wed) as a "discussion section" in which we talk about projects and project preparations, informally. And, then, we could have all the presentations in four meetings held during the final week (6/23–6/26).

- In these lectures, I will layout a general framework for grounding *formal*, *synchronic*, *epistemic coherence requirements* for various types of judgment sets.
- *Formal* coherence is to be distinguished from more substantive types of "coherence" in epistemology [1].
  - Our notions of coherence will (like deductive consistency) supervene on *logical* properties of judgment sets.
- *Synchronic* coherence has to do with the coherence of a set of judgments held by an agent S *at a single time* t.
- *Epistemic* coherence involves *distinctively* epistemic values (specifically: *accuracy* [14] and *evidential support* [8]).
  - Not to be confused with *pragmatic* coherence [25, 13].
- *Coherence* has to do with how a set of judgments "hangs together". CRs are *wide-scope* [2], non-local requirements.
- *Requirements* are *evaluative*; they give *necessary* conditions for (ideal) epistemic rationality of a doxastic state [27].

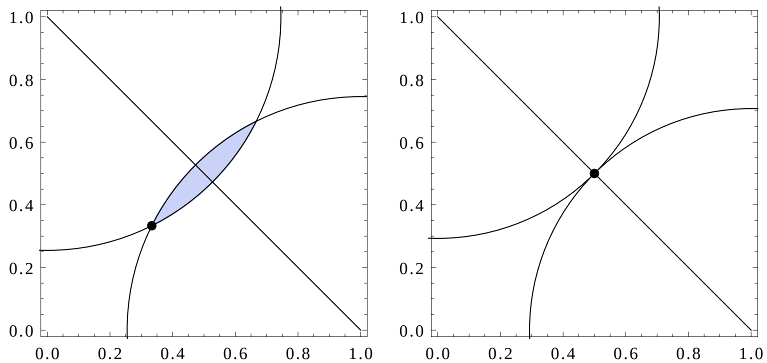
- Our framework is a generalization of Joycean [16, 15] arguments for *probabilism* as a coherence requirement for sets of numerical confidence judgments (*viz*., credences).
- These arguments trace back to de Finetti [9] and they have recently culminated in a vast generalization [24] which forms the basis of our approach to numerical credence.
- Before discussing credences, however, I want to talk about *evidential probabilities*. It will be a basic slogan of our approach that "probabilities reflect evidence". That is:
  - **Probabilities Reflect Evidence** (PRE). For each epistemic context (where a context is determined by a body of total evidence E), there exists a probability function $\Pr_E(\cdot)$ which measures the degree of evidential support in that context.
- In this first lecture, I will explain how Joycean arguments allow us to provide a (partial) justification of (PRE). Then, we will use (PRE) as a key premise in our later arguments.

- Ultimately, our framework will be applicable to any types of judgment which satisfy the following four conditions.
  1. **Alethic Ideal**. The *alethic ideal* for a set of judgments is the conditions under which the set is *perfectly accurate* (in a world). These *supervene on the truth-values* of judged p's.
  2. **Evidential Ideal**. The *evidential ideal* for a set of judgments is the conditions under which the set is *supported by the total evidence* (in a context of epistemic evaluation). Following (PRE), these will have a *probabilistic* explication.
  3. **Non-Factivity**. The judgments in question are not assumed to be factive. That is, one may (sometimes) rationally hold judgments that are inaccurate (*i.e.*, they may *fall short of* 1).
  4. **Fallibility**. The judgments are not assumed to be infallible. One may (sometimes) rationally hold judgments that aren't supported by the total evidence (they may *fall short of* 2).
- *Some* instances of non-factivity + fallibility *will* count as *ir*rational. These will be the *incoherent* doxastic states.

- The basic idea behind our approach traces back to an elegant geometrical argument of de Finetti [9].
- I will present a different interpretation of de Finetti's argument (and the subsequent arguments in the literature).
- On my favored interpretation, these arguments provide (at least) a partial justification for (PRE). They reveal some nice properties of *probabilistic* degrees of evidential support.
- First, some terminology and setup. Let $\mathcal{A} = \{p_1, \ldots, p_n\}$ be a finite *agenda* of n propositions (think of these p_i's as *statements* in a classical propositional language).
- We assume that, in any given epistemic context (with total evidence E), each proposition $p \in \mathcal{A}$ will be supported by E to some (perhaps nil) degree, which we denote as $s(p, E)$.
- Ideally, we'd like to be able to argue that $s(p, E)$ must be a *probability* function. We can use Joyce-style [15] reasoning here. And, Joyce's reasoning traces back to de Finetti's [9].

- I will look at the simplest instance of de Finetti's Theorem, which involves an agenda $\{P, \neg P\}$ for some contingent P.
- Let us suppose (as a matter of convention) that $s(\cdot, E)$ takes values on the unit interval $[0, 1]$, with 0 corresponding to minimal support and 1 corresponding to maximal support.
- Now, let's consider assignments of *credences* (degrees of confidence) to $\{P, \neg P\}$. 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 $b(\neg P)$.
- Next, we assume the following alethic and evidential *ideals*.
  - **Alethic Ideal**. As an *alethic ideal*, $b(\cdot)$ should assign a value of 1 to all (actual) truths and 0 to all (actual) falsehoods.
  - **Evidential Ideal**. As an *evidential ideal*, $b(\cdot)$ should be *equal to* $s(\cdot, E)$, where E is the (actual) total evidence.

- Of course, credence functions which satisfy the evidential ideal will *not* typically satisfy the alethic ideal. The total evidence is *rarely* conclusive (for all p's in any given $\mathcal{A}$).
- But, the evidential ideal should only conflict with the alethic ideal *up to a point*. *Sufficiently radical conflicts* between the alethic and the evidential ideals should be prohibited (or, following the evidence could lead us *too* far astray).
- ☞ If a credence function $b(\cdot)$ is *radically* inaccurate, then it must not reflect the total evidence. The Joycean explication of "radically inaccurate" involves *gradational inaccuracy*.
- The (gradational) inaccuracy of a credal assignment $b(p)$ at a possible world w is some function $\acute{\imath}_b$ of $b(p)$ and the value assigned to p by the *indicator function* $v_w(p) \in [0, 1]$, which assigns 1 (0) to all truths (falsehoods) in w.
- Let $\acute{\imath}_b(b(p), v_w(p)) \overset{\text{def}}{=} (v_w(p) - b(p))^2$. [Note: this yields *Euclidean distance* $\mathcal{I}_b$ between the *vectors* $b(\cdot)$ and $v_w(\cdot)$.]



- **Simplest case of dF's Theorem [9].** 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 [9]). 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.

- More generally, applying our Joycean strategy (to judgment sets $\mathbf{J}$ of type $\mathfrak{J}$) involves going through *Three Steps.*
  - **Step 1**: Identify a precise sense in which *individual* judgments j of type $\mathfrak{J}$ can be (qualitatively) *inaccurate* (or *alethically defective/imperfect*) at a possible world w.
    - In the case of credences, Joyce assumes that $b(p)$ is inaccurate at a possible world w just in case $b(p) \neq v_w(p)$.
  - **Step 2**: Define an *inaccuracy score* $\mathfrak{i}(j, w)$ for individual judgments j of type $\mathfrak{J}$. This is a numerical measure of *how inaccurate* (in the sense of Step 1) j is (at w). For each *set* $\mathbf{J} = \{j_1, \ldots, j_n\}$, we define its *total inaccuracy* at w as the *sum* of the $\mathfrak{i}$-scores of its members: $\mathcal{I}(\mathbf{J}, w) \overset{\text{def}}{=} \sum_i \mathfrak{i}(j_i, w)$.
    - In the case of credences, Joyce assumes that $\mathcal{I}_b(b(\cdot), v_w(\cdot))$ is the *Euclidean distance* between $b(\cdot)$ and $v_w(\cdot)$ on $\mathcal{A}$.
  - **Step 3**: Adopt a *fundamental epistemic principle*, which uses $\mathcal{I}(\mathbf{J}, w)$ to ground a (formal, synchronic, epistemic) coherence requirement for judgment sets $\mathbf{J}$ of type $\mathfrak{J}$.

- We weren't clear on Step 3 in our presentation of de Finetti's argument. The *fundamental epistemic principle* there was:

  Weak Accuracy-Dominance Avoidance (WADA). $b(\cdot)$ should *not be weakly dominated* in inaccuracy (according to $\mathcal{I}_b$). More formally, there should *not* exist a $b'(\cdot)$ (on $\mathcal{A}$) s.t.

  (i) $(\forall w)\,[\mathcal{I}_b(b'(\cdot), v_w(\cdot)) \leq \mathcal{I}_b(b(\cdot), v_w(\cdot))]$, and

  (ii) $(\exists w)\,[\mathcal{I}_b(b'(\cdot), v_w(\cdot)) < \mathcal{I}_b(b(\cdot), v_w(\cdot))]$.[2]

- Weak dominance principles are typically assumed to be requirements of rationality in the context of decision theory (think of $\mathcal{I}_b$ as an alethic measure of *epistemic (dis)utility*).

☞ If (a) we assume our inaccuracy measure $\mathcal{I}_b$ is *Euclidean distance*, and (b) we assume that *violating* (WADA) entails that b is "*radically* inaccurate," then we can conclude that $s(\cdot, E)$ *must be a probability function.* What justifies (a)?

[2] Weak and strict dominance are *equivalent* for Euclidean distance [26].

- We assumed that $\mathcal{I}_b$ was *Euclidean distance.* And, the de Finetti/Joyce argument for probabilism depends sensitively on this choice. Why not some other inaccuracy measure?
- Maher [19] wonders, specifically, why the *Manhattan distance* (*a.k.a.*, the L_1-norm) *isn't* a perfectly good inaccuracy measure (it *doesn't* yield probabilism *via* WADA).
- Joyce [15] gives an interesting "evidentialist" response. The argument concerns a simple agenda involving a 3-sided die.
- Let $P_i \overset{\text{def}}{=}$ a fair, 3-sided die comes up "i". Suppose S has the credence function $b(P_i) = 1/3$. And, suppose S *knows **only** that the die is fair* (*i.e.*, S has *no other* P_i*-relevant evidence*).
- Joyce claims that such an S clearly has "evidentially correct/ideal" credences. Here, Joyce appeals to *The Principal Principle* (PP) [18] to motivate this claim.
- So far, so good. But, *bad news lurks* for Manhattan distance.

- $b'(P_i) = 0$ *strictly dominates* $b(P_i) = 1/3$, according to the Manhattan distance measure of credal inaccuracy.

☞ Assuming that violations of (WADA) constitute "*radically* inaccurate" credences, using the Manhattan distance *rules out* $b(P_i) = 1/3$ as a candidate for $s(P_i, E)$ in this context.

- This seems wrong. $b(P_i) = 1/3$ seems like a good candidate for $s(P_i, E)$ in this context. Moreover, (PP) *entails* this.
- This "evidentialist" reply to Maher leads Joyce to endorse a more general claim, which we call *Probabilistic Admissibility*.

  Probabilistic Admissibility (PA). A credal inaccuracy measure $\mathcal{I}_b$ is *probabilistically admissible* just in case it *fails* to rule out *any probability* function $b(\cdot)$ *via* (WADA).

- Joyce's rationale for (PA) rests on the assumption that

(PRE′) Every probability function will correspond to the chances in some (possible) context. Thus, assuming (PP), *every* Pr-*function describes some possible support function* $s(\cdot, E)$.

- (PA) suffices to disqualify the Manhattan distance, but it not known whether it ensures that (WADA) entails probabilism.
- *Predd et. al.* [24] have given the most general Joyce-style argument to date. They show that (WADA) entails probabilism, provided only that i_b is a *proper scoring rule*.

  Propriety. A measure $i_b(b(p), v_w(p))$ of the inaccuracy of an individual credal judgment $b(p)$ is *proper* just in case every *probabilistic* credence function $b(\cdot)$ minimizes *expected* i_b-inaccuracy, *according to itself*.

- It is easy to show that the Euclidean distance is proper. Let $b(\cdot)$ be a probabilistic credence function. Then, the b-expected value of $i_b(b'(p), v_w(p))$ is given by

$$b(p) \cdot (b'(p) - 1)^2 + (1 - b(p)) \cdot b'(p)^2$$

  which achieves a (unique) minimum at $b'(p) = b(p)$.

☞ Propriety of $\mathcal{I}_b$ ensures that *probabilistic* b*'s never expect other* b'*'s to be more* $\mathcal{I}_b$*-accurate than themselves.*

- It is easy to show that propriety entails probabilistic admissibility (minimization of $\mathcal{I}_b$-expectation entails non-$\mathcal{I}_b$-dominance). But, the converse is an open question.
- *I.e.*, it is open whether propriety can be replaced by the (seemingly) weaker assumption of (PA) in the argument of Predd *et. al.* Pettigrew [23] discusses this conjecture.[3]
- In any case, it seems likely that any inaccuracy measure which runs afoul of (PA) will prove to be inadequate because it rules out some possible evidential support functions.
- If this proves to be right, then we would have a (partial) justification of (PRE), which makes use of the (*prima facie* more plausible) assumption (PRE′) + a (WADA)-based argument for *probabilism, as the evidential ideal for* $b(\cdot)$.

[3]Strictly speaking, we need to assume *truth-directedness* (*i.e.*, that *uniform* decreases in *point-wise* i_b-inaccuracy entail a decrease in *overall* $\mathcal{I}_b$-inaccuracy) in addition to propriety/probabilistic admissibility (see [23]).

- Joyce faces a potential "evidentialist" objection [10].
- Suppose S assigns non-probabilistic credences $b(P) = 0.3$ and $b(\neg P) = 0.3$. Thus, by de Finetti's theorem, b will be $\mathcal{I}_b$-dominated ($\mathcal{I}_b$ = Euclidean distance) by some b''s.
- Interestingly, all of the b''s which $\mathcal{I}_b$-dominate b are such that $b'(P) > 0.3$. This follows from a more general lemma.

  Lemma ([10]). If b' $\mathcal{I}_b$-dominates b, then ***either*** $b(P) > b'(P)$ and $b(\neg P) > b'(\neg P)$ ***or*** $b(P) < b'(P)$ and $b(\neg P) < b'(\neg P)$.

- Now, suppose S learns (exactly) that her total evidence (conclusively) supports the following credal constraint.

  (†) $b(P) \leq 0.3$.

☞ Constraint (†) *rules out* all b''s which $\mathcal{I}_b$-dominate S's b.

- Is being accuracy-dominated by something that your total evidence *rules out* a (*bona fide*, epistemic) *defect* [20]?

- Caie [4] discusses the following self-referential claim.

  (P) S assigns a credence $b(P) < 1/2$.

- Assuming $\mathcal{I}_b$ = Euclidean distance, it seems that the only *non*-$\mathcal{I}_b$-dominated credal assignment S *can adopt* on $\{P, \neg P\}$ is the *non*-probabilistic: $b(P) = 1/2$; $b(\neg P) = 1$!
- This is because some (otherwise) "possible worlds" are *ruled out by the very act of S's adopting certain credences*. For instance, if S assigns $b(P) \geq 1/2$, then P is (made) *false*.
  - See Campbell-Moore's [6, 5] for more on Caie-type cases.
- Caie's problem is an instance of a more general one. We've been assuming a kind of *act/state independence* in our discussion so far. Here, that independence breaks down.
- It can break down for non-semantic reasons (*e.g.*, causal reasons). Greaves [11] and Carr [7] discuss such cases. And, Konek & Levinstein [17] give a possible Joycean response.

- A third worry about Joyce's argument for probabilism questions whether (a) credences *have* an alethic (or accuracy) ideal, and (b) this ideal is that $b(p) = v_w(p)$.
- Re (a), one might worry that judgments of the form $\ulcorner b(p) = r \urcorner$ do not "aim at p's truth" because such judgments *don't represent the world as being such that* p.
  - One can still maintain that it is *epistemically better* to assign higher credence to truths than falsehoods. And, since this *supervenes on the truth-values* (of the p's), it still counts as an *alethic ideal* (if not an "*accuracy* ideal").
- Re (b), even if you agree that $b(\cdot)$ should *match some function* $f(\cdot)$ (on pain of *inaccuracy*), you might reject $f(\cdot) = v_w(\cdot)$. Maybe $f(\cdot)$ = *the chance function* (at w) [12].
  - This seems incorrect to me. The Principal Principle [18] is an *evidential norm* — it tells you what $b(p)$ should match *if S doesn't know the value of* $v_w(p)$. In fact, (PP) implies that if S *does* know $v_w(p)$, then $b(p)$ should match $v_w(p)$.
- See [12, 22, 21, 3] for further discussion of both (a) and (b).

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

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

[3] M. Caie, *Credence in the Image of Chance*, manuscript, 2014.

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

[5] C. Campbell-Moore, *How to Express Self-Referential Probability and Avoid the (Bad) Consequences*, manuscript, 2014.

[6] ______, *Rational Probabilistic Incoherence? A Reply to Michael Caie*, manuscript, 2013.

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

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

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

[10] K. Easwaran and B. Fitelson, *An "Evidentialist" Worry about Joyce's Argument for Probabilism*, *Dialectica*, 2012.

[11] H. Greaves, Epistemic Decision Theory, Mind, 2013.

[12] A. Hájek, *A Puzzle About Degree of Belief*, manuscript, 2010.

[13] ______, *Arguments for — or Against — Probabilism?*, *BJPS*, 2008.

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

[15] 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.

[16] ______, *A Nonpragmatic Vindication of Probabilism*, *Philosophy of Science*, 1998.

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

[18] D. Lewis, *A Subjectivists's Guide to Objective Chance*, 1980.

[19] P. Maher, *Joyce's Argument for Probabilism*, *Philosophy of Science*, 2002.

[20] R. Pettigrew, *Accuracy and Evidence*, *Dialectica*, 2014.

[21] ______, *What Chance-Credence Norms Should Not Be*, *Noûs*, 2013.

[22] ______, *Accuracy, Chance, and the Principal Principle*, *Philosophical Review*, 2012.

[23] ______, *Epistemic Utility Arguments for Probabilism*, *Stanford Encyclopedia of Philosophy*, 2011.

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

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

[26] M. Schervish, T. Seidenfeld, J. Kadane, *Proper Scoring Rules, Dominated Forecasts, and Coherence*, *Decision Analysis*, 2009.

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