A Framework for Grounding Formal Epistemic Coherence Requirements

Branden Fitelson[1]

Department of Philosophy
Rutgers University
&
Munich Center for Mathematical Philosophy
Ludwig-Maximilians-Universität München

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

[1] This presentation draws heavily on joint work with Rachael Briggs (ANU), Fabrizio Cariani (NU), Kenny Easwaran (USC) and David McCarthy (HKU). See http://fitelson.org/coherence/ for further background & results.

- This presentation is about (i) formal, (ii) synchronic, (iii) epistemic (iv) coherence (v) requirements.
  - (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]).
    - Ideally, whether a set of judgments is coherent (in the present sense) should (in principle) be determinable *a priori*.
  - (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* requirements.
  - (iii) *Epistemic* coherence is meant to involve some *distinctively epistemic* values (*e.g.*, *accuracy* and *evidential support*).
    - This is to be distinguished from *pragmatic* coherence (*e.g.*, immunity from arbitrage, dutch books, and the like).
  - (iv) *Coherence* should have something to do with how a set of judgments "hangs together" — in a *wide/global* sense.
  - (v) *Requirements* are *evaluative* principles — for *3rd person assessment* of conformity w/ideals of epistemic rationality.

- Traditional notions of coherence (of the sort we have in mind) have included *deductive consistency* (for full belief) and *probabilism* (for comparative & numerical confidence).
- Lately, there have been interesting discussions about whether there *are* such coherence requirements *at all* [18].
- We will only briefly touch on one of these debates (regarding full belief) here. Today, we'll be content to showcase our approach to grounding such requirements.
- Basically, what we'll be doing is generalizing Joyce's [17, 16] approach to grounding *probabilism* as a coherence requirement for sets of *numerical* confidence judgments.
- ☞ Today, I'll explain how to turn Joyce's argument for probabilism into a *unified framework* for grounding coherence requirements for various types of judgments.
- I'll focus on 3 types of judgments: full belief (B), numerical confidence (b) and comparative/ordinal confidence ($\succeq$).

- At the highest level of abstraction, the framework results in global coherence requirements for sets of judgments $\mathfrak{J}$ (of type J) over (finite) Boolean algebras $\mathcal{B}$ of propositions.
- Applying the framework involves **The Three Steps**:
  - **Step 1**: Define the *vindicated (viz., perfectly accurate) judgment set* (of type J), at world w. Call this set $\mathring{\mathfrak{J}}_w$.
    - Think of $\mathring{\mathfrak{J}}_w$ as the judgments (of type J) that "the *omniscient/ideal* agent" would make (at world w).
  - **Step 2**: Define a notion of "distance between $\mathfrak{J}$ and $\mathring{\mathfrak{J}}_w$". That is, define a measure of *distance from vindication*: $\mathfrak{D}(\mathfrak{J}, \mathring{\mathfrak{J}}_w)$.
    - Think of $\mathfrak{D}$ as measuring how far one's judgment set $\mathfrak{J}$ is (in w) from the vindicated or ideal set of judgments $\mathring{\mathfrak{J}}_w$ (in w).
  - **Step 3**: Adopt a *fundamental epistemic principle*, which *uses* $\mathfrak{D}(\mathfrak{J}, \mathring{\mathfrak{J}}_w)$ to ground a formal coherence requirement for $\mathfrak{J}$.
    - Think of the fundamental epistemic principle as articulating an *evaluative connection* between $\mathfrak{D}$ and $\mathfrak{J}$-coherence.
- This is all way too abstract. Let's look at three examples.

- Before we dive in — some notation and basic assumptions.
  - $B(p) \overset{\text{def}}{=} S$ believes that p. $D(p) \overset{\text{def}}{=} S$ disbelieves that p.

☞ We will be using *simple, idealized models* of epistemic agents (*viz*., epistemically rational agents). As with all models, *only some aspects* of epistemic agents are modeled.

- I'll assume the following about B, D. [The first 4 are integral, but the last one is just for simplicity. More on these later.]
  - **Logical Omniscience**. If $p \dashv\vDash q$, then $B(p) \equiv B(q)$.
  - **Incompatibility**. $B(p) \Rightarrow \neg D(p)$.
  - **Opinionation**. $B(p) \vee D(p)$.
  - **Accuracy conditions**. $B(p)$ [$D(p)$] is accurate iff p is T [F].

  - **(NC)** $D(p) \equiv B(\neg p)$.
- Our S's make judgments regarding *all* p's in some *finite* Boolean algebra $\mathcal{B}$ (generated by some $\mathcal{L}_P$). I will use $\mathfrak{B}$ to denote the set of *all* of S's judgments (B's and D's) over $\mathcal{B}$.

- Classical deductive consistency is implicated in some well-known *paradoxes*. The two most infamous paradoxes (of the sort I have in mind) are the Lottery and the Preface.

  (I) **Lottery Paradox** ([20],[10]). For each ticket i, it is highly probable that i is a loser (L_i). So, it would seem reasonable to be such that $B(L_i)$, for each i. However, this inevitably renders our set $\mathfrak{B}$ *inconsistent*, since we *know* that $(\exists i)(\neg L_i)$.

  (II) **Preface Paradox** ([23],[6]). Let $\mathbf{B} \subset \mathfrak{B}$ be the set containing *all* of your *reasonable* (1^{st}–order) beliefs. This $\mathbf{B}$ is an incredibly rich and complex set of judgments. You're fallible (*i.e.*, your 1^{st}–order evidence is *sometimes misleading*). So, it seems reasonable to believe that *some* B's in $\mathbf{B}$ are false. However, adding *this* (2^{nd}–order) belief to $\mathbf{B}$ renders $\mathfrak{B}$ *inconsistent*.

- There has been a ton of discussion of these paradoxes in the literature. Roughly, there are two sorts of responses:

  (1) Dogmatically *maintain* consistency as a (*global*) CR for $\mathfrak{B}$.
  - Pollock ([24]): *all* inconsistent $\mathfrak{B}$'s engender "*collective defeat*," which mandates *suspension of belief* (wrt *some* p's).

  - Pollock-style responses are implausible. As Christensen [6] nicely explains, it just seems wrong to claim that *suspension* is *reasonable/supported*, for *any* salient p. [More generally, (1) has implausible consequences about *evidential support*.]

  (2) Radically *abandon all* (formal) coherence requirements for $\mathfrak{B}$.
  - Christensen [6] suggests we *do everything with credences*, which (a) *do* have a formal coherence requirement (namely, *probabilism*), and (b) have no trouble dealing with the Lottery or the Preface. We should *learn to live without* B/D.
  - Kolodny [18] thinks we *cannot live without* B/D, but we *should* learn to live without (formal, rational) coherence requirements for them. Instead, they have *only evidential requirements*, which are *satisfied* in the Lottery & Preface.

- I would prefer a different response than *either* (1) *or* (2). But, while (2) goes *too far*, there is a kernel of truth to it.
  - Christensen's focus on credences is useful. We can learn a lot from how *their* coherence requirements are *grounded*.
  - Kolodny's focus on evidence/justification is also crucial.

- We would suggest that these paradoxes indicate that consistency is *too strong* to be a CR for full belief.
- Typically, such paradoxes involve a *conflict* between a *consistency* requirement and an *evidential* requirement, which requires *believing what is justified/supported*.

  ☞ Ideally, we want coherence requirements for full belief that are entailed by *both alethic and evidential* considerations.

- Our framework yields just such CR's, in "3 easy steps".
- **Step 1**: Define the *vindicated (viz., perfectly accurate) judgment set*, at w. ["Judgments of the omniscient S at w."]
  - $\mathring{\mathfrak{B}}_w$ contains $B(p)$ [$D(p)$] iff p is true (false) at w.
- **Step 2**: Define a notion of "distance between $\mathfrak{B}$ and $\mathring{\mathfrak{B}}_w$". That is, a measure of *distance from vindication* $d(\mathfrak{B}, \mathring{\mathfrak{B}}_w)$.
  - $d(\mathfrak{B}, \mathring{\mathfrak{B}}_w) \overset{\text{def}}{=}$ the number of inaccurate judgments in $\mathfrak{B}$ at w.
- **Step 3**: Adopt a *fundamental epistemic principle*, which *uses* $d(\mathfrak{B}, \mathring{\mathfrak{B}}_w)$ to ground a formal coherence requirement for $\mathfrak{B}$.

- Given our choices at Steps 1 and 2, there is ***a*** choice we can make at Step 3 that will yield *consistency* as a CR for $\mathfrak{B}$.

  Possible Vindication (PV). There exists *some* possible world w at which *all* of the judgments in $\mathfrak{B}$ are accurate. Or, to put this more formally in terms of d: $(\exists w)[d(\mathfrak{B}, \mathring{\mathfrak{B}}_w) = 0]$.

- Possible vindication is *one way* we could go here. But, our framework is *much more general* than the classical one. It allows for *many other* choices of fundamental principle.
- Like Joyce [17, 16] — who makes the *analogous* move with *credences*, to ground *probabilism* — we *retreat* from (PV) to the *weaker*: *avoidance of (weak) dominance in* $d(\mathfrak{B}, \mathring{\mathfrak{B}}_w)$.

  Weak Accuracy-Dominance Avoidance (WADA).

  There does *not* exist an alternative belief set $\mathfrak{B}'$ such that:
  (i) $(\forall w)[d(\mathfrak{B}', \mathring{\mathfrak{B}}_w) \le d(\mathfrak{B}, \mathring{\mathfrak{B}}_w)]$, and
  (ii) $(\exists w)[d(\mathfrak{B}', \mathring{\mathfrak{B}}_w) < d(\mathfrak{B}, \mathring{\mathfrak{B}}_w)]$.

- Completing Step 3 in *this* way leads to a *new* CR for $\mathfrak{B}$.

- The new coherence requirement implied by this application of our framework has just the sort of properties we wanted.
- We wanted a coherence requirement that (like consistency) was motivated by considerations of accuracy (ideally, *entailed by* alethic requirements such as consistency/PV).
- But, we also wanted a coherence requirement that was *strictly weaker* than deductive consistency — in such a way that it is *also entailed by our evidential requirements*.
- It can be shown that we have met both of these *desiderata*, provided that we accept the following *necessary condition* for the satisfaction of our evidential requirements (EB).
  - **Necessary Condition for Satisfying (EB)**. S satisfies (EB), *i.e.*, S's judgments are *justified*, ***only if*** **($\mathcal{R}_s$)** there exists *some* probability function $\Pr(\cdot)$ which *probabilifies each of her beliefs and dis-probabilifies each of her disbeliefs*.
- There is *dis*agreement about *which* $\Pr(\cdot)$ is the requisite one for (EB) [5, 27, 14], but there is *agreement* that (EB) $\Rightarrow$ ($\mathcal{R}_s$).

- Here are the logical relationships between key norms:

The Truth Norm for Belief: (TB)
$\Downarrow \; \not\Uparrow$
Possible Vindication (*viz.*, consistency): (PV)
$\Downarrow \; \not\Uparrow$
☞ Weak Accuracy-Dominance Avoidance: $\boxed{\text{(WADA)}}$
$\Uparrow \; \not\Downarrow$
Evidential Requirement for Belief: (EB)

$$\begin{tikzcd}
\text{(TB)} \arrow[d, Rightarrow] & & \text{(EB)} \arrow[ddl, Rightarrow] \\
\text{(PV)} \arrow[dr, Rightarrow] & & \\
& (\mathcal{R}_s) \arrow[d, Rightarrow] \arrow[r, shift left, Rightarrow] & \text{(WADA)} + \text{(NC)} \arrow[l, shift left] \\
& \text{(WADA)} &
\end{tikzcd}$$

- For a more detailed map (with 10 norms), see Extras (28,29).

- (WADA) leads to a coherence requirement that is *strictly weaker* than consistency — *even if* (NC) $D(P) \equiv B(\neg P)$. [We think (WADA) + (NC) is *equivalent* to ($\mathcal{R}_s$). See Extras (27–29).]
- Moreover, in general, *minimal* inconsistent $\mathfrak{B}$'s (of cardinality *greater than 2*) will *not violate* (WADA)/($\mathcal{R}_s$).
- So, if we adopt (WADA)/($\mathcal{R}_s$), rather than (PV)/consistency, "paradoxes" (I) & (II) do *not* imply violations of our CR for $\mathfrak{B}$.
- But, if S *does* violate (WADA)/($\mathcal{R}_s$), this reveals *two* defects:
  - There must exist some judgment in $\mathfrak{B}$ that is *inaccurate*. [This *alethic defect* follows from: (PV) $\Rightarrow$ (WADA)/($\mathcal{R}_s$).]
  - *And*, there must exist some judgment in $\mathfrak{B}$ that is *not supported by "the evidence"* — *whatever* "the evidence" is! [This *evidential defect* follows from: (EB) $\Rightarrow$ (WADA)/($\mathcal{R}_s$).]

☞ Our coherence requirements are *points of agreement* between alethic *vs* evidential perspectives (*i.e.*, points of contact between truth/accuracy *vs* justification/evidence).

- As always, the application of our framework will involve our *three steps*. In the case of *sets of credences* $\mathfrak{b}$, this means:
  - **Step 1**: Define "the *vindicated credal set at w*" ($\mathring{\mathfrak{b}}_w$).
    - There will be *greater controversy* about $\mathring{\mathfrak{b}}_w$ than $\mathring{\mathfrak{B}}_w$. But, in the end, I think the best explication of $\mathring{\mathfrak{b}}_w$ is quite clear.
  - **Step 2**: Define "the distance between $\mathfrak{b}$ and $\mathring{\mathfrak{b}}_w$" [$\delta(\mathfrak{b}, \mathring{\mathfrak{b}}_w)$].
    - Much of the extant literature involves this (choice of δ) step. We'll see that there is not as much robustness/agreement here as there was in the analogous step for full belief.
  - **Step 3**: Choose a *fundamental principle* that *uses* $\delta(\mathfrak{b}, \mathring{\mathfrak{b}}_w)$ to ground a coherence requirement for for credal sets $\mathfrak{b}$.
    - In this context, the choice(s) of fundamental principle are (in some ways) even more straightforward than in the case of full belief. [The $\mathfrak{b}$-analogue of (PV) is a *non-starter* here.]
- Since the application to credal sets was the first (historical) example of the framework we're describing, there is an extant literature on these Three Steps, and their perils.

- **Step 1**: define "the *vindicated* credal set at w" ($\mathring{\mathfrak{b}}_w$).
- Let $v_w(\cdot)$ be the 0/1–*truth-value-assignment* associated with w. That is, $v_w(p) = 1$ iff p is T at w and $v_w(p) = 0$ iff p is F at w. A simple way to state Joyce's [17] definition of $\mathring{\mathfrak{b}}_w$ is:

$$\mathring{\mathfrak{b}}_w \stackrel{\text{def}}{=} \{b(p) = r \mid v_w(p) = r\}$$

- So, on the Joycean approach, $\mathring{\mathfrak{b}}_w$ is *the set of extremal credence assignments corresponding to the 0/1-truth-value assignments* associated with world w. This suggests:

$$\frac{p \text{ is true}}{B(p)} :: \frac{v_w(p) = r}{b(p) = r}$$

- Of course, this analogy is a poor one, *if* it is interpreted as articulating a ***norm** that actual agents are meant to **follow***.
- But, this is only meant as an *evaluative* analogy, for the *omniscient/ideal* agent. In this sense, the analogy is sound.
  - Some [15] have proposed $\mathring{\mathfrak{b}}_w \stackrel{\text{def}}{=}$ *the set of values given by the chance function at w* [ch_w]. This is an inappropriate *def*.

- **Step 2**: define "the *distance* from $\mathfrak{b}$ to $\mathring{\mathfrak{b}}_w$" [$\delta(\mathfrak{b}, \mathring{\mathfrak{b}}_w)$].
- As in the case of full belief, this second step is fraught with potential danger/objections. *Many* δ's are possible here.
- Moreover, unlike the case of full belief, there is *strong disagreement* here — *even between naïve candidate* δ's.
- The norms we end-up with (assuming analogous choices of fundamental principles in Step 3, below) will *depend sensitively* on which distance measure δ is chosen.
- Let's start by thinking about what sorts of mathematical representations of $\mathfrak{b}$'s are most natural. [In the case of opinionated $\mathfrak{B}$, *binary vectors* were the natural choice.]
- In this case, it is natural to represent $\mathfrak{b}$'s as *vectors in* $\mathbb{R}^n$, where n is the number of propositions in the underlying $\mathcal{B}$.
- So, the natural things to consider are *measures of distance between vectors in* $\mathbb{R}^n$. For a nice survey, see: [9, Ch. 5].

- I'll focus on two natural (l_p*–metric* [9, Ch. 5]) choices for δ
  - $\delta_1(\mathfrak{b}, \mathring{\mathfrak{b}}_w) \stackrel{\text{def}}{=} \sum_p |b(p) - v_w(p)|$
  - $\delta_2(\mathfrak{b}, \mathring{\mathfrak{b}}_w) \stackrel{\text{def}}{=} \sqrt{\sum_p |b(p) - v_w(p)|^2}$
- δ_1 (the l_1–metric) is also called *Manhattan distance*, and δ_2 (the l_2–metric) is, of course, the *Euclidean distance*.
- Interestingly, these two natural choices of δ will lead to drastically different CR's for $\mathfrak{b}$ in our framework [22].
- de Finetti [8] proved the following (show "geometric proof")
  - **Theorem** (de Finetti). A credal set $\mathfrak{b}$ is *strictly* δ_2-dominated by some alternative credal set $\mathfrak{b}'$ *iff* $\mathfrak{b}$ is *non-probabilistic*.
- *I.e.,* being *non*-probabilistic is *equivalent* to *violating* a *strict* δ_2-dominance avoidance requirement [call this (SADA_{δ_2})].
  - It can be shown [26] that the (SADA_{δ_2}) and (WADA_{δ_2}) renditions of de Finetti's theorem are *equivalent*.

- Before discussing the (radical) differences between the CR's yielded by δ_1 *vs* δ_2 [(WADA$_{\delta_2}$) *vs* (WADA$_{\delta_1}$)], let's think about **Step 3** in the contexts of full belief *vs.* credence.
- For full belief, (PV) is *too strong*. But, this can only be seen clearly by considering "paradoxical" cases involving *large-ish*, *minimal-inconsistent* $\mathfrak{B}$'s (more on *that* tomorrow).
- In such cases, it is unreasonable to require consistency, as this is in tension with our *evidential* requirements.
- In the credal context, (PV) seems ***patently*** *too strong*. For it would require one's credences to be *extremal* (always).
- Of course, it often seems *clear* to us — *even in very simple, non-paradoxical examples* — that having extremal credences would be *unreasonable* (*viz.*, *un*supported by our evidence).
- It is for this reason that fundamental principles weaker than (PV) are *even more attractive* in the credal case.

- The story *changes drastically* if we move from δ_2 to δ_1.
  - There are *non-probabilistic* credence sets $\mathfrak{b}$ that are *not even weakly dominated* in δ_1-distance from vindication. In our toy example, suppose $\mathfrak{b}$ contains $b(P) = b(\neg P) = 0$. This credal set $\mathfrak{b}$ is *not weakly δ_1-dominated* by ***any*** $\mathbb{R}^2$-vector.
  - It is still true, in $\mathbb{R}^2$, that *probabilistic* credal sets will *also* be *non*-dominated (even weakly), assuming measure δ_1. [This is *trivial* in the $\mathbb{R}^2$-case, since *no* $\mathbb{R}^2$-vector from $[0,1]$ is weakly δ_1-dominated by *any other* $\mathbb{R}^2$-vector from $[0,1]$!]
  - In $\mathbb{R}^3$, the story about δ_1 becomes even more interesting.
    - $\langle 0, 0, \frac{1}{2}\rangle$ weakly (but *not strictly*) δ_1-dominates $\langle \frac{1}{4}, \frac{1}{4}, \frac{1}{2}\rangle$.
    - $\langle 0, 0, \frac{1}{2}\rangle$ *strictly* δ_1-dominates $\langle \frac{3}{16}, \frac{3}{16}, \frac{5}{8}\rangle$.

☞ Therefore, in the general case, *neither direction* of de Finetti's theorem carries over from δ_2 to δ_1!

  - Notice how there *is* a difference between weak and strict δ_1-dominance. This is *not so* for δ_2. Generally, δ's that make de Finetti's theorem *true* imply no such distinction [26].

- Let ⌜$p \succeq q$⌝ be the binary relation (relative to an agent S at time t, *viz.*, over a finite Boolean Algebra $\mathcal{B}$) ⌜S is at least as confident in the truth of p as she is in the truth of q (at t)⌝.
- There is widespread ([7], [19], [25], [11]), though not perfect ([12]), agreement that these "intuitive" axioms govern $\succeq$.

(A1) $(p \succeq q) \vee (q \succeq p)$.

(A2) If $p \succeq q$ and $q \succeq r$, then $p \succeq r$.

(A3) $\top \succ \bot$. [*i.e.*, $(\top \succeq \bot)$ & $(\bot \not\succeq \top)$]

(A4) $p \succeq \bot$.

(A5) If $\langle p, q\rangle$ and $\langle p, r\rangle$ are mutually exclusive, then:

$$q \succeq r \iff (p \vee q) \succeq (p \vee r).$$

- I will explain how one can give a *direct*, *accuracy-dominance* justification for (most of) these axioms in our framework.
- As always, this will involve completing "the three steps."

- Let $\mathfrak{C}$ be the set of all of S's comparative $\succeq$-judgments (over all pairs in $\mathcal{B} \times \mathcal{B}$). Step 1 defines "*the vindicated set at w*" ($\mathring{\mathfrak{C}}_w$). And, Step 2 defines "distance from $\mathfrak{C}$ to $\mathring{\mathfrak{C}}_w$" [$\delta(\mathfrak{C}, \mathring{\mathfrak{C}}_w)$].
- Our first attempt(s) to complete Steps 1 and 2 for $\succeq$ failed. The mistake we made was trying to *locally score* each pairwise judgment in $\mathfrak{C}$ *as if it were akin to a full belief* [21].
- Rather than explain our initial (failed) attempt, I'll describe our *new way* (which *works* & fits better into our framework).
- Here is a plausible fundamental comparative principle:

(T$\succeq$) Vindication requires being *strictly more confident* in (all) truths than (all) falsehoods (*i.e.*, if $v_w(p) > v_w(q)$, then $p \succ q$ should be included in the w-vindicated relation: $\mathring{\mathfrak{C}}_w$).

- This implies *one condition* for inclusion in the set $\mathring{\mathfrak{C}}_w$. But, (T$\succeq$) *leaves open* various comparisons at various worlds.
- Specifically, (T$\succeq$) is *silent* on comparisons between pairs of propositions that *have the same truth-value at w*.

- One is tempted to find a (single) constraint on these remaining pairs of propositions on which (T$\succeq$) is silent.
- And, the *natural* proposal along these lines is:

  ($\mathcal{I}$) If $v_w(p) = v_w(q)$, then $p \sim q$ should be included in $\mathring{\mathfrak{C}}_w$.

- If we were to add ($\mathcal{I}$) to our basic principle (T$\succeq$), then we would end-up with the following *definition* of $\mathring{\mathfrak{C}}_w$.
  - $\mathring{\mathfrak{C}}_w$ is the set containing $p \succ q$ ($p \sim q$) iff p is true at w and q is false at w (p and q have the same truth-value at w).
- In other words, on this proposal, $\mathring{\mathfrak{C}}_w$ would correspond to *the ordering* $\succeq_w$ *that is* (i) indifferent between all truths in w, (ii) indifferent between all falsehoods in w, and (iii) ranks all truths in w *strictly above* all falsehoods in w.
- This does not seem like a crazy initial proposal. After all, if we *were* to assume that there *are* numerical credences around, then $\succeq_w$ would just be *the relation (uniquely) represented by the vindicated credence function* $\mathring{\mathfrak{b}}_w$.

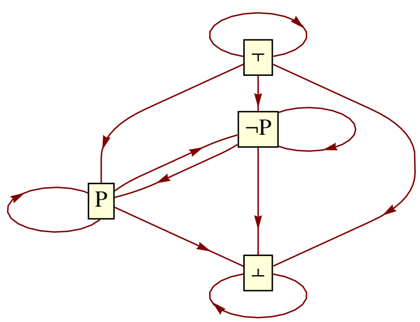
- *Technically*, this (simpler) proposal is (nearly) *equivalent* to a (more complex) approach that we favor, *philosophically* [as it does *not* assume $p \sim q$ is *vindicated* when $v_w(p \equiv q) = 1$].
- A binary relation $\succeq$ on $\mathcal{B}$ (containing n propositions) can be represented by its $n \times n$ *binary adjacency matrix*.
- So, a natural way to think of distance between binary relations on $\mathcal{B}$ is as distance between adjacency matrices. But, this is just *distance between binary vectors of length* n^2.
- We already have our preferred, näive way of measuring *that*: *Hamming distance*. And, that will be our Step 2 proposal.
- This corresponds to the most oft-used distance measure between binary relations: the *Kemeny distance* [9, §10.2].
- To illustrate how the Kemeny distance works (and how our simple definition of $\mathring{\mathfrak{C}}_w/\succeq_w$ works), I will apply Kemeny distance to a couple of simple/toy orderings and $\succeq_w$'s.

- Suppose we adopt $\succeq_w$ as the definition of *the vindicated comparative confidence relation at* w; and, we measure distance between relations as Kemeny distance. That is:

$\delta(\mathfrak{C}, \mathring{\mathfrak{C}}_w) \overset{\text{def}}{=}$ Hamming distance between the adjacency matrices of S's comparative confidence relation $\succeq$ and the relation $\succeq_w$.

- Consider a toy agent S with the following $\succeq$ relation ($\mathfrak{C}$).

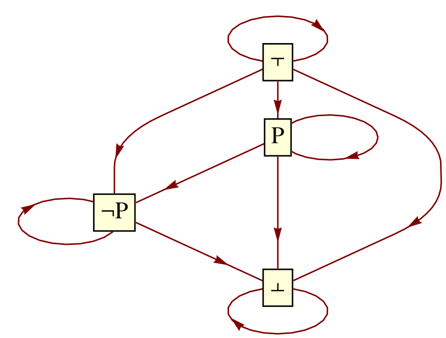
$\mathfrak{C}$	$\top$	P	$\neg P$	$\bot$
$\top$	1	1	1	1
P	0	1	1	1
$\neg P$	0	1	1	1
$\bot$	0	0	0	1



- Suppose $P \overset{\text{def}}{=}$ a fair coin lands heads; and, S knows ***only*** that the coin is fair (*i.e.*, S has *no other* P*-relevant evidence*).

- Intuitively, this agent's relation *shouldn't be ruled-out as incoherent* (it seems to accord with evidential requiements).
- Happily, if we adopt Kemeny distance (δ) as our distance measure, then $\mathfrak{C}$ will *not* be *weakly* δ*-dominated by* ***any*** $\mathfrak{C}'$.
- Of course, I won't show this *by hand*. It is easily verified by computer. However, in order to illustrate how δ works (and what $\mathring{\mathfrak{C}}_w/\succeq_w$ look like), consider the following *specific* $\mathfrak{C}'$:

$\mathfrak{C}'$	$\top$	P	$\neg P$	$\bot$
$\top$	1	1	1	1
P	0	1	1	1
$\neg P$	0	0	1	1
$\bot$	0	0	0	1



- Here is what $\succeq_w$ looks like for w_1 and w_2 in our toy case:

$\succeq_{w_1}$	$\top$	P	$\neg P$	$\bot$
$\top$	1	1	1	1
P	1	1	1	1
$\neg P$	0	0	1	1
$\bot$	0	0	1	1

$\succeq_{w_2}$	$\top$	P	$\neg P$	$\bot$
$\top$	1	1	1	1
P	0	1	0	1
$\neg P$	1	1	1	1
$\bot$	0	1	0	1

- And, again, here is what $\mathfrak{C}$ and $\mathfrak{C}'$ look like:

$\mathfrak{C}$	$\top$	P	$\neg P$	$\bot$
$\top$	1	1	1	1
P	0	1	1	1
$\neg P$	0	1	1	1
$\bot$	0	0	0	1

$\mathfrak{C}'$	$\top$	P	$\neg P$	$\bot$
$\top$	1	1	1	1
P	0	1	1	1
$\neg P$	0	0	1	1
$\bot$	0	0	0	1

- These δ's reveal that $\mathfrak{C}'$ does *not* (even weakly) dominate $\mathfrak{C}$:
  - $\delta(\mathfrak{C}, \succeq_{w_1}) = 3 > 2 = \delta(\mathfrak{C}', \succeq_{w_1})$.
  - $\delta(\mathfrak{C}, \succeq_{w_2}) = 3 < 4 = \delta(\mathfrak{C}', \succeq_{w_2})$.

- I won't get into details here. But, it can be shown [13] that:

  Theorem. If an agent's $\succeq$-relation $\mathfrak{C}$ exhibits any violation (N) of any of the "intuitive" axioms (A1)–(A5) for $\succeq$, then there exists a relation $\mathfrak{C}'$ that (a) weakly δ-dominates $\mathfrak{C}$, and (b) does not exhibit (N). [Think of $\mathfrak{C}'$ as a "local fix" of (N).]

- Moreover, in some cases [(A1), (A3), (A4)], we can also show that the $\mathfrak{C}'$ in question is *itself not* δ-dominated by *any* relation defined over *just* the propositions implicated in (N).
- For the other two axioms [(A2), (A5)], we can show that (provided $\mathfrak{C}$ is total and $\sim$-transitive) if $\mathfrak{C}$ *instantiates A* (with respect to a *particular set* of propositions $\mathbf{P}$), then there will *not* exist a (total) relation $\mathfrak{C}'$, which (a) agrees with $\mathfrak{C}$ on everything *outside* of $\mathbf{P}$, and (b) weakly δ-dominates $\mathfrak{C}$.
- We are increasingly optimistic that the following is true:

  Conjecture. Provided that $\mathfrak{C}$ is total and $\sim$-transitive, $\mathfrak{C}$ is *not* weakly δ-dominated by *any* (total) relation $\mathfrak{C}'$ ***iff*** $\mathfrak{C}$ is representable by some numerical probability function.

	$\mathfrak{B}_1$	$\mathfrak{B}_2$	$\mathfrak{B}_3$	$\mathfrak{B}_4$	$\mathfrak{B}_5$	$\mathfrak{B}_6$	$\mathfrak{B}_7$	$\mathfrak{B}_8$
$\neg X \,\&\, \neg Y$	D	D	D	D	D	D	D	D
$X \,\&\, \neg Y$	D	D	D	D	D	D	D	D
$X \,\&\, Y$	D	D	D	D	D	D	D	D
$\neg X \,\&\, Y$	D	D	D	D	D	D	D	D
$\neg Y$	D	D	D	D	B	B	B	B
$X \equiv Y$	D	D	B	B	D	D	B	B
$\neg X$	D	B	D	B	D	B	D	B
X	B	D	B	D	B	D	B	D
$\neg(X \equiv Y)$	B	B	D	D	B	B	D	D
Y	B	B	B	B	D	D	D	D
$X \vee \neg Y$	B	B	B	B	B	B	B	B
$\neg X \vee \neg Y$	B	B	B	B	B	B	B	B
$\neg X \vee Y$	B	B	B	B	B	B	B	B
$X \vee Y$	B	B	B	B	B	B	B	B
$X \vee \neg X$	B	B	B	B	B	B	B	B
$X \,\&\, \neg X$	D	D	D	D	D	D	D	D

- There are judgment sets that are *non-d-dominated, but inconsistent*.
- This is true, *even if* we assume *opinionation* & (NC) $D(p) \equiv B(\neg p)$.
- The simplest possible examples of this kind involve $\mathcal{L}$'s with *at least two atomic sentences* (*e.g.*, X, Y).
- The table shows *the* 8 sets that are *non-dominated*, but *inconsistent*.
- Of course, as we look at larger and larger languages/algebras, we start seeing *more interesting* examples.
- But, even these toy examples exhibit a "preface-like" structure.
- The *logically stronger* propositions tend to be *dis*believed (and the *logically weaker* ones *believed*).

(TB) S ought believe p just in case p is true.

(PV) $(\exists w)[d(\mathfrak{B}, \mathring{\mathfrak{B}}_w) = 0]$. That is, $\mathfrak{B}$ is *deductively consistent*.

(SADA) $\nexists \mathfrak{B}'$ such that: $(\forall w)[d(\mathfrak{B}', \mathring{\mathfrak{B}}_w) < d(\mathfrak{B}, \mathring{\mathfrak{B}}_w)]$.

$\neg(\exists_\beta \Box$MI) $\nexists \beta \subseteq \mathfrak{B}$ s.t.: $(\forall w)\left[> \frac{1}{2} \textit{ of the members of } \beta \textit{ are inaccurate at } w\right]$.

($\mathcal{R}_s$) $\exists$ a probability function $\Pr(\cdot)$ such that, $\forall p \in \mathcal{B}$:

$$B(p) \textit{ iff } \Pr(p) > \tfrac{1}{2}, \text{ and } D(p) \textit{ iff } \Pr(p) < \tfrac{1}{2}.$$

(EB) S ought believe p just in case p is supported by S's evidence.

Note: this assumes *only* $(\exists \Pr)(\forall p)\left[\Pr(p) > \frac{1}{2} \textit{ iff } B(p)\right]$.

$\neg(\exists_\beta \Box$HI$\Diamond$MI) $\nexists \beta \subseteq \mathfrak{B}$ s.t.: $(\forall w)\left[\geq \frac{1}{2} \textit{ of the members of } \beta \textit{ are inaccurate at } w\right]$ & $(\exists w)\left[> \frac{1}{2} \textit{ of the members of } \beta \textit{ are inaccurate at } w\right]$.

(WADA) $\nexists \mathfrak{B}'$ s.t.: $(\forall w)[d(\mathfrak{B}', \mathring{\mathfrak{B}}_w) \leq d(\mathfrak{B}, \mathring{\mathfrak{B}}_w)] \,\&\, (\exists w)[d(\mathfrak{B}', \mathring{\mathfrak{B}}_w) < d(\mathfrak{B}, \mathring{\mathfrak{B}}_w)]$.

$\neg(\exists_\beta \Box$HI) $\nexists \beta \subseteq \mathfrak{B}$ s.t.: $(\forall w)\left[\geq \frac{1}{2} \textit{ of the members of } \beta \textit{ are inaccurate at } w\right]$.

(NC) S disbelieves p iff S believes $\neg p$ [*i.e.*, $D(p) \equiv B(\neg p)$].

- Here is what the logical relations look like, among all of the 10 norms for (opinionated) $\mathfrak{B}$. [Double (single) arrows represent *known* (*conjectured*) entailments. And, if there is no path, then we believe (or conjecture) that there is no entailment.]

$$
\begin{tikzcd}
 & (TB) \arrow[d, Rightarrow] & & (EB) \arrow[ddl, Rightarrow] & \\
 & (PV) \arrow[dr, Rightarrow] & & & \\
\neg(\exists_\beta \Box HI) \arrow[rr, shift left] \arrow[d, Rightarrow] & & (\mathcal{R}_s) \arrow[ll, Rightarrow, shift left] \arrow[rr, Rightarrow, shift right] \arrow[d, Rightarrow] & & (WADA) + (NC) \arrow[ll, shift right] \\
\neg(\exists_\beta \Box HI \Diamond MI) \arrow[rr, Leftrightarrow] \arrow[d, Rightarrow] & & (WADA) \arrow[d, Rightarrow] & & \\
\neg(\exists_\beta \Box MI) \arrow[rr, Leftrightarrow] & & (SADA) & &
\end{tikzcd}
$$

- We (along with Rachael Briggs and Fabrizio Cariani) [4] are investigating various applications of this new approach.
- One interesting application is to *judgment aggregation*. *E.g.*,
  - Majority rule aggregations of the judgments of a bunch of agents — each of whom satisfy (PV) — *need not* satisfy (PV).
- **Q**: does majority rule preserve *our* notion of coherence, *viz.*, is (WADA) preserved by MR? **A**: yes (on simple, atomic + truth-functional agendas), but *not on all possible agendas*.
  - There are (not merely atomic + truth-functional) agendas A and sets of judges J ($|A| \geq 5$, $|J| \geq 5$) that (severally) satisfy (WADA), while their majority profile *violates* (WADA).
- *But*, if a set of judges is (severally) *consistent* [*i.e.*, satisfy (PV)], then their majority profile *must* satisfy (WADA).

☞ **Recipe.** Wherever $\mathfrak{B}$-*consistency* runs into paradox, substitute *coherence* (in *our* sense), and see what happens.

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

[2] K. Easwaran and B. Fitelson, *Accuracy, Coherence & Evidence*, in progress, 2012.

[3] ______, *An "Evidentialist" Worry about Joyce's Argument for Probabilism*, *Dialectica*, 2012.

[4] R. Briggs, F. Cariani, K. Easwaran and B. Fitelson *Individual Coherence and Group Coherence*, manuscript, 2012.

[5] R. Carnap, *Logical Foundations of Probability*, U. of Chicago, 2nd ed., 1962.

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

[7] B. de Finetti, *Foresight: Its Logical Laws, Its Subjective Sources*, in H. Kyburg and H. Smokler (*eds.*), *Studies in Subjective Probability*, Wiley, 1964.

[8] ______, *The Theory of Probability*, Wiley, 1974.

[9] M. Deza and E. Deza, *Encyclopedia of Distances*, Springer, 2009.

[10] I. Douven and T. Williamson, *Generalizing the Lottery Paradox*, *BJPS*, 2006.

[11] T. Fine, *Theories of Probability*, Academic Press, 1973.

[12] P. Fishburn, *The Axioms of Subjective Probability*, *Statistical Science*, 1986.

[13] B. Fitelson and D. McCarthy, *Probabilism Without Numbers*, in progress, 2012.

[14] R. Fumerton, *Metaepistemology and Skepticism*, Rowman & Littlefield, 1995.

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

[16] 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.

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

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

[19] B. Koopman, *The axioms and algebra of intuitive probability*, *Ann. of Math.*, 1940.

[20] H. Kyburg, *Probability and the Logic of Belief*, Wesleyan, 1961.

[21] B. Levinstein, *Fit and Comparative Probability: Dim Prospects for de Finetti*, 2012.

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

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

[24] J. Pollock, *The Paradox of the Preface*, *Philosophy of Science*, 1986.

[25] L. Savage, *The Foundations of Statistics*, Dover, 1972.

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

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