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.

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 \( b(H) = b(T) = \frac{1}{2} \) 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 \( i_b \) of \( b(p) \) and \( w \)’s *indicator function* \( v_w(p) \), which assigns 1 (0) to truths (falsehoods) in \( w \). The overall inaccuracy \( I_b \) of \( b(\cdot) \) at \( w \) is \( I_b(b, w) = \sum_p i_b(b(p), v_w(p)) \).

• de Finetti’s measure \( i_p(b(p), v_w(p)) \leq (v_w(p) - b(p))^2 \) is known as the Brier Score. This implies \( I_b \) 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* \( b’(\cdot) \)’s (on \( \{P, \neg P\} \)). The point \((1, 0)\) \((0, 1)\) corresponds to the world in which \( P \) is true (false).

Theorem (de Finetti [8]). \( b \) is non-probabilistic \( \iff \exists b’(\cdot) \) which is (Euclidean) closer to \( v_w(\cdot) \) in every possible world.

• The plot on the left (right) explains the \( \Rightarrow \) \((=)\) 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) \( p \) 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 \( A \) 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 \( B \) (on \( A \)).

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 \((r)\) with making an accurate judgment, and some negative utility \((-w)\) with making an inaccurate judgment (where \( w > r > 0 \)).

Because suspensions are neither accurate nor inaccurate \((\text{per se})\), 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 \( u(\cdot, w) \).

\[
\begin{align*}
    u(B(p), w) & \equiv \begin{cases} 
        -w & \text{if } p \text{ is false at } w \\
        r & \text{if } p \text{ is true at } w 
    \end{cases} \\
    u(D(p), w) & \equiv \begin{cases} 
        r & \text{if } p \text{ is false at } w \\
        -w & \text{if } p \text{ is true at } w 
    \end{cases} \\
    u(S(p), w) & \equiv \begin{cases} 
        0 & \text{if } p \text{ is false at } w \\
        0 & \text{if } p \text{ is true at } w 
    \end{cases}
\end{align*}
\]

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 (synchronous) 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 \( B \) over an agenda \( A \) should, from the point of view of their own credence function \( b(\cdot) \), maximize expected epistemic utility (or minimize expected inaccuracy). That is, \( B \) should maximize

\[
EEU(B, b) \equiv \sum_{p \in A} \sum_{w \in W} b(w) \cdot u(B(p), w)
\]

where \( B(p) \) is the agent’s attitude toward \( p \), and \( W \equiv \bigcup A \).

We also assume “act-state independence”: \( B(p) \) 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 \( B \) over agenda \( \mathcal{A} \) satisfies

**Coherence** if and only if for all \( p \in \mathcal{A} \)

\[
\begin{align*}
B(p) \in B & \iff b(p) > \frac{w}{r + w}, \\
D(p) \in B & \iff b(p) < \frac{r}{r + w}, \\
S(p) \in B & \iff b(p) \in \left[ \frac{r}{r + w}, \frac{w}{r + w} \right].
\end{align*}
\]

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.**

<table>
<thead>
<tr>
<th>Epistemic Principle</th>
<th>Analogous Decision-Theoretic Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alethic Ideal</strong></td>
<td>(AMU) Do ( \phi ) only if ( \phi ) maximizes utility in the actual world.</td>
</tr>
<tr>
<td><strong>Consistency</strong></td>
<td>(PMU) Do ( \phi ) only if ( \phi ) maximizes utility in some possible world.</td>
</tr>
<tr>
<td><strong>Coherence</strong></td>
<td>(MEU) Do ( \phi ) only if ( \phi ) maximizes expected utility (relative to some ( Pr )).</td>
</tr>
<tr>
<td><strong>(WADA)</strong></td>
<td>(WDOM) Do ( \phi ) only if ( \phi ) is not weakly dominated in utility.</td>
</tr>
<tr>
<td><strong>(SADA)</strong></td>
<td>(SDOM) Do ( \phi ) only if ( \phi ) is not strictly dominated in utility.</td>
</tr>
</tbody>
</table>

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 \( B_n \) 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 \( B_n \) (with threshold \( t \)) entails deductive consistency of \( B_n \iff t \geq \frac{n-1}{n} \).

If we combine this with Easwaran’s **Coherence** theorem, we get the following result, regarding the conditions under which the **Coherence** of \( B_n \) entails the consistency of \( B_n \).

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

\[
\omega \geq (n - 1) \cdot r.
\]

Insisting that **Coherence implies consistency** (wrt \( B_n \)) requires (naïve) agents to disvalue inaccuracy at least \( (n - 1) \) times as much as they value accuracy.
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 $B'$ such that:

(i) $\forall w \in \{B', w\} \leq u(B, w)$, and

(ii) $\exists w \in \{B', w\} < u(B, w)$.

**Strict Accuracy-Dominance Avoidance (SADA).**

There does not exist an alternative belief set $B'$ such that:

(iii) $\exists w \in \{B', w\} < u(B, w)$.

It turns out [10, 11] that Coherence $\Rightarrow$ (WADA) $\Rightarrow$ (SADA).

Indeed, (WADA) and (SADA) are very weak [10]. But, they do constitute non-trivial necessary requirements of rationality.

I mentioned above that we assume “act-state independence” (ASI). There are two main reasons we assume (ASI) here.

If $B(p)$ 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 $B(p)$ 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].

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 $\perp$ is a logical contradiction, then it would not be rational for $S$ to believe $\perp$ (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 $\langle p \& q \rangle$ at $t$.

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

(SPC) If $p \equiv q$, then any $B$ s.t. $\{B(p), D(q)\} \subseteq B$ is incoherent.

(NCB) Any $B$ such that $\{B(\perp)\} \subseteq B$ is incoherent.

$\neg$(CP) Not every $B$ s.t. $\{B(p), B(q), D(p \& q)\} \subseteq B$ is incoherent.

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 $P$. [$\neg B(\langle P \rangle).$]

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.

<table>
<thead>
<tr>
<th>$P$</th>
<th>$\neg P$</th>
<th>$B(P)$</th>
<th>$B(\neg P)$</th>
<th>$D(P)$</th>
<th>$D(\neg P)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>$F$</td>
<td>$T$</td>
<td>$-$</td>
<td>$+$</td>
<td>$-$</td>
</tr>
<tr>
<td>$w_2$</td>
<td>$T$</td>
<td>$F$</td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
</tbody>
</table>

The “$x$”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)?
We (along with Rachael Briggs and Fabrizio Cariani) [2] are investigating various applications of the approach in [10].

One interesting application is to judgment aggregation. E.g.,

- Majority rule aggregations of the judgments of a group of consistent agents need not be consistent.

Q: does majority rule preserve our notion(s) of coherence, e.g.,, 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 (or merely Coherent), then their majority profile must be Coherent.

Recipe. Wherever B-consistency runs into paradox, substitute coherence (in our sense), and see what happens.

In the Meno (97e–98a), Socrates says:

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.

- Our epistemic utility function (for belief) only assigned positive value to correctness. What about knowledge?

- Nothing in our (teleological) framework for epistemic utility theory rules out attaching (additional) value to knowledge — over and above the value we place on correctness.

- Julien Dutant and I are investigating such models. They can be used to ground knowledge-based Lockeans theses. E.g.,

  One ought to believe that $p$ when one judges that it is (sufficiently) highly probable that one knows that $p$.

---

*References*