Probabilistically coherent credences despite opacity

Christian List[*]

Abstract: Real human agents, even when they are rational by everyday standards, sometimes assign different credences to objectively equivalent statements, such as "Orwell is a writer" and "E.A. Blair is a writer", or credences less than 1 to necessarily true statements, such as not-yet-proven theorems of arithmetic. Anna Mahtani calls this the phenomenon of "opacity". Opaque credences seem probabilistically incoherent, which goes against a key modelling assumption of probability theory. I sketch a modelling strategy for capturing opaque credence assignments without abandoning probabilistic coherence. I draw on ideas from judgment-aggregation theory, where we face similar challenges of defining the "objects of judgment".

Anna Mahtani (2024) argues that credence claims are "opaque" in the sense that someone may assign different credences (degrees of belief, subjective probabilities) to statements with equivalent propositional content (a phenomenon of "hyperintensionality"). At first sight, this goes against a key modelling assumption in probability theory. Ordinarily, the objects of probability assignments – whether objective or subjective – are propositions (sometimes also called "events"), and any two statements that express the same proposition (describe the same event) must be assigned the same probability. But as Mahtani notes, real people's beliefs, when expressed as real-valued credences, often violate this constraint, even when these people are rational by everyday standards. Mahtani investigates how such opaque credence assignments can be accommodated within a credence framework.

In this short paper, I sketch what I consider an attractive modelling strategy for capturing opaque credence assignments without overthrowing some of the core features of the probability-theoretic framework. My proposal is broadly in line with one of the proposals Mahtani discusses (namely, that "states", for the purposes of the credence framework, could be something other than metaphysically possible worlds), but it draws on modelling techniques from judgment-aggregation theory, where we face a similar challenge of defining the "objects of judgment".

---

1. The credence framework

Credences, or *degrees of belief*, are classically modelled as (subjective) probability functions on some set of propositions. The formalism has the following ingredients.

Propositions or *events*, which are different labels for the same formal concept, are modelled as subsets of some underlying set Ω of "possible worlds" or "possible states". Any *proposition* (or *event*) is identified with the set of worlds or states at which the proposition is true or the event is taking place. A *probability function* then assigns to each proposition or event a real number between 0 and 1, which can be interpreted as the (subjective or objective) probability of that proposition or event.

More formally, let Ω be some non-empty set. We usually call the elements of Ω "worlds" or "states", but from a formal perspective, they could be anything. In the formalism, they are simply treated as mutually exclusive and jointly exhaustive of some relevant space of "possibilities". Any subset of Ω is called a *proposition* or *event*. We can think of the intersection of two propositions/events p and q, namely $p \cap q$, as their *conjunction*, their union $p \cup q$ as their *disjunction*, and the complement of any proposition/event p, namely $\Omega \setminus p$, as its *negation*.

An *algebra* is a non-empty collection $\mathcal{A}$ of propositions/events that is closed under union, intersection, and complementation. A simple way to generate an algebra is to let $\mathcal{A}$ be the set of all subsets of Ω, but sometimes there may be other canonical ways of defining an algebra $\mathcal{A}$ based on Ω. A *probability function* Pr is a function from a given algebra $\mathcal{A}$ into the interval [0,1] with two properties:

- $\Pr(\Omega) = 1$
- $\Pr(p \cup q) = \Pr(p) + \Pr(q)$ whenever p and q are disjoint (i.e., their intersection is empty).

In sum, credences are classically represented by probability functions on some algebra of propositions or events and interpreted as representing an agent's degrees of belief on the elements of that algebra. For the rest of this paper, I will follow the convention in philosophy of speaking of "propositions" rather than "events", but nothing of substance hinges on that terminological choice.

2. The opacity problem

An algebra offers a simple calculus for propositions, allowing us to represent conjunctions, disjunctions, and negations of propositions, and probability assignments are constrained by these logical relations. However, an algebra doesn't have the expressive resources of a language (whether formal or natural), and this also restricts the expressive resources of the credence framework. For instance:

- While in a language (even in a simple one such as propositional logic) there can be distinct sentences that are equivalent in semantic content, the propositions in an algebra can never be distinct yet equivalent in semantic content.
- While in a language there can be more than one tautological (or necessarily true) sentence and more than one contradictory (or necessarily false) one, any algebra contains only a single tautology (represented by the entire set Ω) and a single contradiction (represented by the empty set $\emptyset$).

These limitations of an algebra are perfectly fine for the modelling purposes of probability theory in mathematics. However, if we want to use the probability-theoretic apparatus for representing a real human agent's credal state – thereby re-interpreting the framework as a kind of psychological theory – we run into difficulties.

As Mahtani notes, an agent may assign different credences to statements with equivalent propositional content, contrary to what the standard probability-theoretic formalism would seem to require. Here are two examples (the first of which is discussed by Mahtani herself):

Example 1: George Orwell and Eric Arthur Blair are, as a matter of fact, the same person (Orwell was just the pen name of Blair), and yet an agent might have a credence of 0.8 that Orwell is a writer but only a credence of 0.1 that Blair is a writer (not knowing that Orwell is Blair). The probability-theoretic formalism doesn't seem to be able to capture this insofar as the sentences "Orwell is a writer" and "Blair is a writer" pick out the same proposition.

Example 2: A mathematician might assign a credence of 1 to Fermat's last theorem, which was proved by Andrew Wiles, but a lower credence to Goldbach's conjecture ("every even number greater than 2 is the sum of two primes"), which has not yet been proved, although the latter, like the former, might turn out to be a necessary truth of arithmetic. It seems that the probability-theoretic formalism would mandate the same credence assignment (of 1) in each case, insofar as two distinct necessarily true sentences would each correspond to the same tautological proposition (Ω).

As Mahtani notes, the solution to this problem is not obvious, and there are several distinct stategies one might pursue. We could, for example, make sentences (from either natural language or some formal language) the objects of credence, but it then becomes difficult to formulate conditions on credence assignments which preserve the idea that one's credence assignments for some sentences constrain those that one can rationally make for others while still capturing the intended opacity phenomena. If we required that *the credence assigned to a sentence must always be the credence assigned to its propositional content*, then we would fail

to capture the intended phenomena; we would be back where we started. By contrast, if we relaxed the italicized requirement, then we would have to explain what other constraints, if any, to impose upon credence assignments across different sentences in order to avoid the implication that "anything goes".

For example, we may plausibly want to say that it is a rationality violation

(1) to assign different credences to "Orwell is a writer" and "Orwell is an author",

while it is rationally permissible

(2) to assign different credences to "Orwell is a writer" and "Blair is a writer".

Why might (1) be rationally impermissible while (2) is permissible? In the sentence pair in (1), the equivalence in meaning is transparent, while in the sentence pair in (2) it is not. However, it's a challenge to come up with a "sentential" account of the objects of credence that neatly captures this difference without running into all sorts of technical difficulties.

Mathani notes that if, instead, we wish to retain a framework in which the objects of credence are formally the elements of some algebra (rather than sentences in some language), we may need to drop the common assumption (common in philosophy, I should emphasize) that the set Ω of worlds or states from which that algebra is derived must consist of metaphysically possible worlds. She explores the route of dropping that assumption and allowing states to be "something else". Mahtani writes (2024, ch. 8):

> "One natural proposal along these lines is to take states to be sets of sentences. On this proposal, there are many details to be filled in and challenges to be met. What language or languages are the sentences in? How does context fit into this picture? Should the worlds be complete and/or coherent, and in what sense? How these questions are answered will have numerous repercussions for users of the credence framework."

In what follows, I want to pursue a version of this modelling strategy, albeit a particularly abstract (but hopefully also very flexible) one, building on parallel challenges we have faced in judgment-aggregation theory. I don't want to claim much originality here, however, since the proposal is not only in line with some of Mahtani's ideas, but it also echoes ideas that we find in philosophical discussions of impossible worlds (e.g., Berto and Jago 2019, Pettigrew 2021).

3. My proposal

In judgment-aggregation theory, we study how groups of individuals can make collective judgments on some issues based on the individuals' judgments on those issues (List and Pettit

2002, Dietrich 2007, List and Puppe 2009). The individuals whose judgments are to be aggregated could be judges in a collegial court, experts on some expert panel, legislators in a parliament, members of another assembly or committee, or simply citizens in a democracy. The issues could be whether a defendant is liable for breach of contract, how atmospheric greenhouse gases affect the global climate, which policies will promote economic growth, or what the future vision for a country should be.

A key modelling question is: how should we formally represent the "issues" on which judgments are made, i.e., the objects of judgment? Should we think of them as

- sentences from some language (either natural or formal),
- propositions understood as subsets of some set of possible worlds,
- elements of some algebra, or
- simply abstract "yes"/"no" questions (the answers to which can be expressed as k-tuples of zeros and ones)?

Different modelling choices (reviewed in Dietrich and List forthcoming) have their pros and cons, just as different ways of modelling the objects of credence have their pros and cons, as discussed by Mahtani (2024). Inspired by judgment-aggregation theory, I want to sketch a modelling strategy, which – when applied to credences – allows us to retain a definition of credences as coherent probability assignments to the elements of some algebra, and yet to capture the opacity phenomena of interest (Examples 1 and 2 above). The strategy involves:

- first defining an abstract "language" for modelling purposes,
- then using this to generate a suitable algebra of propositions, and
- finally defining credences as probability assignments on this algebra.

We begin by defining a *language* in a very thin and abstract sense (following Dietrich 2007), namely as a non-empty set **L** whose elements we call "sentences" (we could also think of them as "statements"), which is endowed with

- a *negation operator* $\neg$ such that for any sentence p in **L**, there is another sentence $\neg p$ in **L**, called its *negation*, and
- a notion of *consistency*, which I will here call *(subjective) tenability*, which partitions the set of all subsets of **L** into those that are *tenable* and those that aren't.

To give a plausible example, a set of the form $\{p, q, p \wedge q\}$ might be tenable, where p and q are mutually logically independent sentences and $p \wedge q$ is their conjunction, while a set of the form $\{p, \neg p\}$ is untenable. For simplicity, we assume that double-negations cancel each other out. To be well-behaved for modelling purposes, the notion of tenability (we could alternatively call it "subjective consistency") must satisfy four conditions:

- *Negation-untenability*: Sentence-negation pairs are untenable.
- *Monotonicity*: Subsets of tenable sets are tenable.
- *Non-triviality*: The empty set is tenable.
- *Completability*: Any tenable set has a tenable superset containing a member of each sentence-negation pair in **L**.

This definition is very flexible and permissive. The set **L** could be pretty much anything. Although we call its elements "sentences" (or "statements"), there is no assumption about whether they are natural-language sentences, formal-language sentences, or of any other form, as long as there is a negation operator. They could be numbers, strings of symbols, or any objects that can be elements of a set. The notion of tenability can also be defined and interpreted in any way we like. The only restriction is that the four stated conditions must be satisfied.

Indeed, beginning with a set **L** of sentences that has a negation operator, one can easily engineer a notion of tenability in line with these conditions via the following recipe:

- Consider the set of all subsets of **L** that contain precisely one member of each sentence-negation pair in **L**.
- Stipulate which of these sets are *tenable* (and which not); these are then deemed to be the *maximal tenable* subsets of **L** (those which have no tenable superset in **L**).
- Deem any subset of **L** *tenable* if and only if it is a subset of one of these maximal tenable sets.

Even if the elements of **L** are originally drawn from some natural or formal language, such as English or the language of arithmetic, tenability in **L** need not be defined in the same way in which consistency or satisfiability would be defined for the background natural or formal language. Tenability in **L** could be far more permissive. In arithmetic, for instance, the set consisting of Peano's axioms and the negation of Goldbach's conjecture might be objectively inconsistent, if Goldbach's conjecture turned out to be a true theorem, but in **L** such a set could be subjectively tenable.

A natural interpretation of tenability in **L** is tenability from the perspective of the (non-omniscient and perhaps boundedly rational) agent whose judgments or beliefs we wish to model, not tenability from the perspective of an omniscient Olympian observer. So, as the terminology suggests, it makes sense to think of tenability in **L** as subjective rather than objective tenability. Any set of sentences from **L** whose inconsistency or mutual incompatibility an agent is not – or not currently – able to establish could be deemed tenable.

We can then easily define languages **L** in which sets such as the following are tenable:

- {"Orwell is a writer", "Blair is not a writer"} or

- the set consisting of Peano's axioms and the negation of some complicated theorem of arithmetic.

In **L**, someone who accepts such sets of sentences would not breach any (subjective) tenability constraints. The notion of (subjective) tenability in **L** further induces a notion of (subjective) entailment: a set S of sentences in **L** *(subjectively) entails* another sentence p in **L** if and only if $S \cup \{\neg p\}$ is not subjectively tenable in **L**. Just as fewer things may be untenable in **L** than in some background language, so fewer things entail one another in **L**. For example, from the agent's subjective perspective, Peano's axioms need not entail Goldbach's conjecture even if from some objective perspective – outside **L** – they might.

So far, we have defined a language for modelling purposes in which more sets of sentences may be tenable than what some objective background notion of consistency might say, and fewer sentences may be deemed equivalent or to stand in entailment relations to one another. The sorts of sentences to which an agent might assign different credences even though they are objectively equivalent could be examples of this.

To model the assignment of credences explicitly, we need to define a suitable algebra. We begin by using **L** to construct a set Ω of "worlds" as follows (List 2019). Let Ω be the set of all maximal tenable subsets of **L**. The elements of Ω formally behave like "worlds". By the completability condition, each maximal tenable subset of **L** contains precisely one member of each sentence-negation pair in **L**. Each element of Ω can thus be interpreted as encoding a subjectively admissible assignment of truth-values ("true", "false") to all the sentences in **L**. Furthermore, the elements of Ω are mutually exclusive and jointly exhaustive of some space of subjectively tenable "possibilities". The set Ω can thus be interpreted as consisting of "personally possible worlds" in the sense discussed by Pettigrew (2021). Pettigrew offers the following interpretation (p. 9995):

> "Roughly speaking, a world is personally possible for a particular individual at a particular time if by this time this individual hasn't ruled it out by their experiences, their logical reasoning, their conceptual thinking, their insights, their emotional reactions, or whatever other cognitive activities and processes can rule out worlds for an individual."

We can further think of subsets of Ω as *(subjective) propositions*, and we can think of the sentences in **L** as expressing such propositions. The *(subjective) propositional content* of any sentence p in **L** *according to this model* will be given by the subset of Ω consisting precisely of those maximal tenable subsets of **L** (elements of Ω) that contain that sentence.

We can now introduce an algebra $\mathcal{A}$ of such propositions. In the simplest case, it could be the set of all subsets of Ω, but it could also be smaller than that. Credences can be defined as probability assignments on that algebra, formally as a function Cr from $\mathcal{A}$ into the interval $[0,1]$, and they can satisfy the standard conditions of probabilistic coherence:

- $\text{Cr}(\Omega) = 1$
- $\text{Cr}(p \cup q) = \text{Cr}(p) + \text{Cr}(q)$ whenever p and q are disjoint.

The credence for any sentence in **L** will then be the credence of its propositional content according to the present construction. Provided the relevant proposition (i.e., the set of maximal tenable subsets of **L** containing the sentence) is in the algebra $\mathcal{A}$ (which it will be, for instance, if $\mathcal{A}$ is the set of all subsets of Ω), then the credence for such a sentence will be well-defined.

Nonetheless, our way of constructing **L** ensures that two sentences that are objectively equivalent from some external perspective, such as "Orwell is a writer" and "Blair is a writer", need not be subjectively equivalent in **L**, and by implication, they may have a different propositional content in $\mathcal{A}$. Therefore, the assignment of different credences to them doesn't breach any constraints of probabilistic coherence with respect to this algebra.

Similarly, in a suitably constructed language **L**, the negation of Goldbach's conjecture may be subjectively tenable together with the Peano axioms or together with the affirmation of other theorems of arithmetic such as Fermat's last theorem, and so the assignment of a credence of less than 1 to Goldbach's conjecture need not breach any constraints of probabilistic coherence.

It is worth noting that our model need not ascribe to the agent unrealistically rich credence assignments, taking the agent to have credences for all sorts of things far removed from his or her cognitive consideration. The set **L** could itself be quite small, possibly just consisting of a small number of sentences that are under cognitive consideration by the agent, so the induced set Ω and any resulting algebra $\mathcal{A}$ could also be quite small. Moreover, a credence function Cr need not be defined on all of $\mathcal{A}$ but could be defined on a much smaller, negation-closed subset X of $\mathcal{A}$ (an "agenda" as defined in judgment-aggregation theory), and probabilistic coherence of Cr would be the requirement of extendability to a full probability function Pr on $\mathcal{A}$.

A possible objection to the modelling strategy I have sketched is that it appears to support a certain kind of subjectivism about possibility and even propositional content. After all, my analysis relies on attributing to each agent a subjective notion of tenability, which then induces a subjective space of "possibilities" for that agent (his or her "personally possible worlds") and gives rise to a subjective account of any sentence's "propositional content". An objector might worry that, on this account, we lack a good way of saying when two people recognize the same possibilities or when one recognizes more possibilities than the other.

My response is that my modelling strategy does offer some resources for comparing the possibilities recognized by different agents. Specifically, when we hold the set **L** of sentences fixed (or consider a set **L** that is the intersection of the sets of sentences on which all the agents in question form their attitudes), we can compare different notions of tenability for **L** by constructing a partial ordering between them. One notion of tenability, say tenability$_1$, is *at least as permissive* as another, say tenability$_2$, if and only if any subset of **L** that is tenable$_2$ is also tenable$_1$. Someone with a more permissive notion of tenability will recognize as tenable anything deemed tenable by someone with a less permissive notion, though not vice versa. We can then also relate the sets of "personally possible worlds" of different agents to one another. Someone's set of personally possible worlds (each of which is represented by a maximally tenable subset of **L** for that agent) could be a subset (or conversely, a superset) of someone else's. In short, we can partially order different agents' sets of personally possible worlds by inclusion. Whenever the partial ordering thus constructed allows comparisons, we can unproblematically compare the possibilities recognized by different agents.

Furthermore, even when two or more agents' notions of tenability cannot be compared in this way, we can construct a "disjunctive" notion of tenability by deeming any subset of **L** *tenable in the disjunctive sense* if and only if it is tenable according to at least one of the relevant agents' notions of tenability. (Similarly, one might also construct a "conjunctive" notion of tenability, though this is perhaps less useful.) It can easily be verified that the disjunctive tenability notion satisfies the four well-behavedness conditions on a notion of tenability (assuming the individual agents' notions do), and the induced set of "possible" worlds (given by the maximally tenable subsets of **L** according to the disjunctive notion of tenability) will be the union of the different agents' personally possible worlds. This union should in principle be rich enough for the analysis of some shared discourse among those agents. Finally, we could study the dynamics of an agent's notion of tenability. Through a learning process, this notion could become less permissive than it was before, and the agent's set of personally possible worlds could shrink.

Needless to say, the details of all this are beyond the scope of this paper. I simply mention those points as avenues for responding to the given objection.

4. Concluding remarks

I have sketched a modelling strategy – broadly within the category for which Mahtani uses the label "states as something else" – which can accommodate several of the opacity phenomena that motivate Mahtani's book, while still defining credences as coherent probability assignments on some algebra. This suggests that agents can have credences that are rational by their own lights while displaying the sorts of opacity/hyperintensionality phenomena that Mahtani investigates.

That said, although this addresses one shortcoming of the credence framework, I don't wish to commit myself to the claim that the credence framework is the right framework for modelling the beliefs of real people. I am still drawn to the view that the credence framework is useful for some modelling purposes, but that, from a psychological or cognitive-science perspective, human beliefs are better modelled differently, for instance as qualitative beliefs, but with various degrees of entrenchment or with epistemic operators expressing degrees of certainty, or in some other non-probabilistic way. But that's a separate issue from the one discussed here.

Author bio: Christian List is Professor of Philosophy and Decision Theory at LMU Munich and Co-Director of the Munich Center for Mathematical Philosophy. He is also a Visiting Professor in Philosophy at the London School of Economics. He works at the intersection of philosophy, economics, and political science, with a particular focus on individual and collective decision-making and the nature of intentional agency.

References

Berto, Francesco, and Mark Jago. 2019. *Impossible Worlds.* Oxford: Oxford University Press.

Dietrich, Franz. 2007. A Generalised Model of Judgment Aggregation. *Social Choice and Welfare* 28(4): 529–565.

Dietrich, Franz, and Christian List. Forthcoming. *Judgment Aggregation*. Book manuscript available on request.

List, Christian. 2019. Levels: Descriptive, Explanatory, and Ontological. *Noûs* 53(4): 852–883.

List, Christian, and Philip Pettit. 2002. Aggregating Sets of Judgments: An Impossibility Result. *Economics and Philosophy* 18(1): 89–110.

List, Christian, and Clemens Puppe. 2009. Judgment Aggregation: A Survey. In *Oxford Handbook of Rational and Social Choice*, ed. P. Anand, C. Puppe, and P. Pattanaik, 457–82. Oxford: Oxford University Press.

Mahtani, Anna. 2024. *The Objects of Credence*. Oxford: Oxford University Press.

Pettigrew, Richard. 2021. Logical ignorance and logical learning. *Synthese* 198(10): 9991–10020.