Trust the Evidence: Two Deference Principles for Imprecise Probabilities

Giacomo Molinari
giacomo.molinari@bristol.ac.uk

July 14, 2023

University of Bristol

Starting intuition

Rational agents value the evidence.

Q: how is this intuition captured by our probabilistic models of belief?

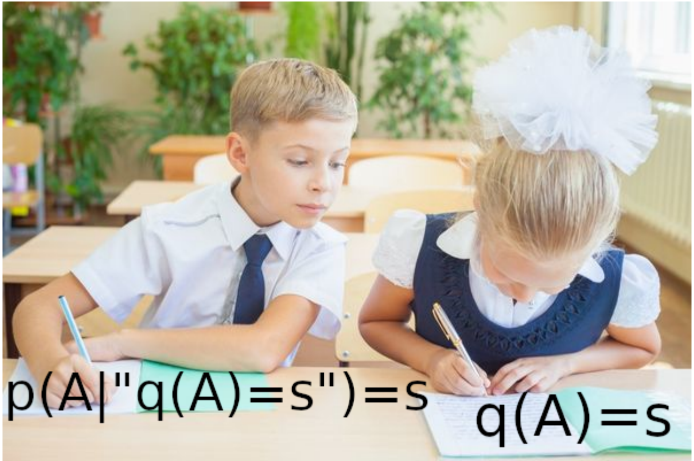
Deference Characterisation

- **Deference Characterisation**: Rational agents value the evidence because they defer to their informed selves, i.e. they treat their updated credal state as an expert.
- **Question**: What does it mean to treat a credal state as an expert?

Deference Principles

- **Deference principles** formally specify the relationship that must hold between two credal sets for one to treat the other as an expert.
- They allow us to specify interesting constraints on a credal set. E.g. “you should defer to your doctor’s opinion”.

Value Reflection: Intuitive Specification



You defer to me only if, conditional on me assigning a certain probability to an event, you assign the same probability to that event.

Definite Descriptions

- Let $\Omega = \{\omega_1, ..., \omega_n\}$.
- A **definite description** of a credal set is a function R that associates to each possibility ω_i a credal set R_i.

Example

- $\Omega = \{\omega_1, \omega_2\}$, $\omega_1 =$ Biden wins, $\omega_2 =$ Trump wins.
- E.g. $R =$ "The (current) credal set of the next US president".
- $R_1 =$ Joe Biden's current credal set.
- $R_2 =$ Donald Trump's current credal set.

- For $S \subseteq \mathbb{R}$, let $[R(A) = S]$ be the event that S is the set of probability values R assigns to A, i.e. $\{\omega_i : R_i(A) = S\}$.
- E.g. $[R(\{\omega_2\}) = \{1\}]$ is the event that the next US president is (currently) fully confident Trump will will the election.

Value Reflection: Formal Definition

- Let Π be your credal set.

Value Reflection

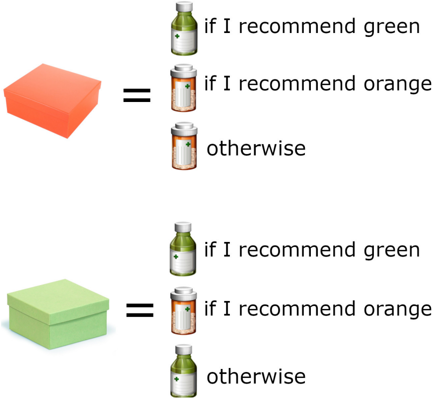
You defer to R iff for every event $A \subseteq \Omega$ and value set $S \subseteq \mathbb{R}$:

$$\Pi(A|[R(A) = S]) = S \tag{1}$$

whenever this conditional credal set is defined.

- **Fact**: all coherent **precise** agents defer to their updated credences under Value Reflection.
- **Problem**: some coherent **IP** agents do not defer to their updated credences under Value Reflection.

Total Trust: Intuitive Specification



You defer to me only if you prefer (don't disprefer) whatever is in the orange box to the orange medicine, and you prefer (don't disprefer) whatever is in the green box to the green medicine.

Total Trust: Formal Definition

- Let Π be your (regular) credal set.
- For any credal set P, let $D_P = \{X : p(X) > 0 \text{ for all } p \in P\}$
- Let $[X \in D_R]$ be the event that R finds X strictly desirable, i.e. $\{\omega_i : X \in D_{R_i}\}$.

Strong Total Trust (S-Trust)

You defer to R iff for every gamble $X : \Omega \to \mathbb{R}$:

$$X \in D_{\Pi(\cdot|[X \in D_R])} \tag{2}$$

whenever this conditional credal set is defined.

Weak Total Trust (W-Trust)

You defer to R iff for every gamble $X : \Omega \to \mathbb{R}$:

$$-X \notin D_{\Pi(\cdot|[X \in D_R])} \tag{3}$$

whenever this conditional credal set is defined.

Deference to Informed Self

Proposition

Let Π be a regular credal set, $\mathcal{E} = \{E_1, ..., E_n\}$ a partition, and R the credal set obtained by updating Π on whichever E_i is true. Then Π S-Trusts R.

- So agents with coherent imprecise credences defer to their informed selves, in the sense of S-Trust.
- This is a way in which they value the evidence.

Thanks!

See you at the poster session!

Trust the Evidence: Two Deference Principles for Imprecise Probabilities

Giacomo Molinari giacomo.molinari@bristol.ac.uk
Department of Philosophy, University of Bristol

University of BRISTOL

The Value of Evidence

- We like to think rational agents value learning new evidence. Here is one way to capture this intuition:

Good's Theorem
For an agent with coherent precise credences, it is never admissible to pay in order to avoid learning free evidence in a sequential decision problem [1].

- **Problem**: For agents with coherent imprecise credences, it is sometimes admissible to pay in order to avoid learning free evidence in a sequential decision problems [2, 3].
- But sequential decision theory comes with a number of philosophical questions (e.g. what counts as an available action?) [4]. Whether Good's theorem holds for IP agents depends on how we answer these questions.
- **Question**: Can we capture the intuition that rational agents value the evidence without appealing to sequential decision theory?
  - **Value of Evidence - Deference (VE-D)**: An agent values the evidence when, for any partition $\mathcal{E}$, she **defers** to her credences updated on the true $E_i \in \mathcal{E}$.

Deference Principles and Imprecise Probabilities

- Let Π be a regular credal set. Let R be the definite description of a credal set, meaning that R may denote a different credal set R_i depending on which $\omega_i \in \Omega$ is the case. We use π and p instead of Π and R when both credal sets are singletons.
- **Deference Principles** specify the relationship that must hold between Π and R for the former to treat the latter as an expert.
- **Fact**: For any partition $\mathcal{E}$, an agent with coherent precise credences *defers* to her credences updated on whichever $E_i \in \mathcal{E}$ is true when deference is defined by **Precise Reflection**.

Precise Reflection

π defers to p iff, for every event $A \subseteq \Omega$ and $s \in \mathbb{R}$:

$$\pi(A | [p(A) = s]) = s \quad (1)$$

whenever this conditional probability is defined, where $[p(A) = s] = \{\omega_i : p_i(A) = s\}$.

- Here is a natural generalisation of Precise Reflection to the imprecise case:

Value Reflection

Π defers to R iff for every event $A \subseteq \Omega$ and value set $S \subseteq \mathbb{R}$:

$$\Pi(A | [R(A) = S]) = S \quad (2)$$

whenever this conditional credal set is defined, where $[R(A) = S] = \{\omega_i : R_i(A) = S\}$

- **Problem**: Sometimes IP agents do not defer to their updated credences according to Value Reflection. There are cases where you are sure that you will have $R(A) = (0, 1)$ after learning the true element of a partition, but you have $\Pi(A) \neq (0, 1)$ before learning [5].

Objective

Find a notion of deference for IP that satisfies the following desiderata:

1. (D1) It captures intuitions about what it means to defer to an expert.
2. (D2) For any partition $\mathcal{E}$, an agent with coherent IP credences defers to their credences updated on the true $E_i \in \mathcal{E}$.
3. (D3) It collapses to a reasonable precise deference principle when all credences involved are precise.

Two Deference Principles for Imprecise Probabilities

- For any credal set P, let $D_P = \{X : p(X) > 0 \text{ for every } p \in P\}$ its corresponding set of desirable gambles.

Strong Total Trust (S-Trust)

Π defers to R iff for every gamble $X : \Omega \to \mathbb{R}$:

$$X \in D_{\Pi(\cdot | [X \in D_R])} \quad (3)$$

whenever this conditional credal set is defined, where $[X \in D_R] = \{\omega_i \in \Omega : X \in D_{R_i}\}$.

Weak Total Trust (W-Trust)

Π defers to R iff for every gamble $X : \Omega \to \mathbb{R}$:

$$-X \notin D_{\Pi(\cdot | [X \in D_R])} \quad (4)$$

whenever this conditional credal set is defined, where $[X \in D_R] = \{\omega_i \in \Omega : X \in D_{R_i}\}$.

D1: Capture Deference Intuition

- Let $\mathcal{A} = \{a_1, a_2\}$ a binary decision problem. The *black-box option* s_i is equivalent to R's preferred option in $\mathcal{A}$, if R has a strict preference in $\mathcal{A}$, and is equivalent to a_i otherwise.
- **Intuition**: if you consider R to be an expert, you should prefer/not disprefer the black-box option s_i to a_i.

Proposition 1
Π S-Trusts R iff for every problem $\mathcal{A} = \{a_1, a_2\}$:

1. If $[s_1 \neq a_1] \neq \emptyset$, then Π strictly prefers s_1 to a_1,
2. if $[s_2 \neq a_2] \neq \emptyset$, then Π strictly prefers s_2 to a_2.

Proposition 2
Π W-Trusts R iff for every problem $\mathcal{A} = \{a_1, a_2\}$:

1. Π does not strictly prefer a_1 to s_1,
2. Π does not strictly prefer a_2 to s_2.

D2: Defer to Informed Self

- Coherent IP agents value the evidence by deferring to their informed selves (VE-D).

Proposition 3
Let Π be a regular credal set, $\mathcal{E} = \{E_1, ..., E_k\}$ be a partition such that $\Pi(\cdot | E_s)$ is defined for every $E_s \in \mathcal{E}$, and denote by R the credal set obtained by updating Π on whichever $E_s \in \mathcal{E}$ is true. Then Π S-Trusts R.

D3: Reasonable Precise Restriction

- If both Π and R are singletons, then both STT and WTT are equivalent to the following precise deference principle, defended in [6]:

Total Trust

π defers to p iff for every gamble X:

$$\pi(X | [p(X) \geq 0]) \geq 0 \quad (5)$$

whenever this conditional prevision is defined, where $[p(X) \geq 0] = \{\omega_i : p_i(X) \geq 0\}$

Open Questions

- Can W/S-Trust be modified to produce interesting constraints, of the kind expressed by Propositions 1 and 2, for arbitrary decision problems?
- Can we extend Total Trust to an IP deference principle that is sensitive to differences between credal sets which are not reflected in their sets of desirable gambles?

References

[1] Irving John Good. On the principle of total evidence. *British Journal for the Philosophy of Science*, 17(4):319–321, 1967.
[2] Seamus Bradley and Katie Steele. Can free evidence be bad? value of information for the imprecise probabilist. *Philosophy of Science*, 83(1):1–28, 2016.
[3] Joseph B Kadane, Mark Schervish, and Teddy Seidenfeld. Is ignorance bliss? *The Journal of Philosophy*, 105(1):5–36, 2008.
[4] Jason Konek. Predicting coherence. Forthcoming.
[5] Roger White. Evidential symmetry and mushy credence. *Oxford studies in epistemology*, 3:161–186, 2010.
[6] Kevin Dorst, Benjamin A Levinstein, Bernhard Salow, Brooke E Husic, and Branden Fitelson. Deference done better. *Philosophical Perspectives*, 35(1):99–150, 2021.