Trust the evidence; deference principles for imprecise probability

Giacomo Molinari

March 29, 2023

1 Introduction

We like to think that rational agents *value evidence*. Precise Bayesians commonly capture this intuition by appealing to Good's theorem (Good, 1967). The theorem shows that, if a rational agent is offered the option to learn some new evidence for free before facing a decision problem, they are not willing to pay to turn down this offer. In other words: rational agents never pay to avoid free evidence.

Good's theorem seems to fail if we allow agents to have imprecise credences. Rational imprecise agents are allowed to pay in order to avoid free evidence, and depending on the imprecise decision theory one picks, they may even be required to do so (Bradley and Steele, 2016; Kadane et al., 2008). This raises the worry that the coherence and decision-making norms of imprecise probability are not norms of rationality: agents who follow them don't value evidence, and are therefore not rational.

In this essay I will respond to this worry by arguing that our starting intuition that rational agents value evidence is better captured by appealing to a *deference principle*, rather than in terms of sequential decision problems. Deference principles tell us, given an agent's credal state, which credal states they consider as experts. The intuition that rational agents value evidence should be captured as the claim that rational agents defer to their informed selves, treating their informed selves like an epistemic authority. I will introduce and defend two deference principles for imprecise credences, and show that imprecise agents defer to their informed selves according to these principles under the assumptions of Good's theorem. This shows a way in which imprecise agents can be said to value the evidence.[1] I will end by showing that even in the precise setting, there are cases where Good's theorem does not apply, but where the requirement that one should defer to one's informed self, as captured by appropriate deference principles, imposes substantive constraints on a rational agent's credences. This further supports the view that our intuitions about the value of free evidence are best captured in terms of deference.

[1] See (Dorst, 2020) for a discussion of how Good's theorem relates to deference principles in the precise case.

2 The Value of Evidence

2.1 Some notation

A *finite probability space* is a pair (Ω, p) where Ω is a finite possibility space, and $p : 2^\Omega \to \mathbb{R}$ is a probability function. I will assume throughout that the possibility space Ω is finite, and will speak of a probability function instead of a probability space whenever there is no risk of confusion regarding the function's domain. I use $\mathcal{P}_\Omega$ to describe the set of all probability functions on a given Ω.

We model an agent's individual probabilistic judgements as sets of probability functions. For example, let H be the event that a coin lands heads, and T the event that it lands tails. Then the judgement that the coin is fair can be captured by the set $\{p \in \mathcal{P}_\Omega : p(H) = 1/2\}$ of probability functions which assign probability $1/2$ to H. The judgement that the coin is biased towards heads can be captured by the set $\{p \in \mathcal{P}_\Omega : p(H) > 1/2\}$ of all functions which assign greater probability to H than to T.

We model an agent's entire doxastic state by a single nonempty set $P \subseteq \mathcal{P}_\Omega$, known as the agent's *credal set*. The idea is that an agent makes a probabilistic judgement, such as the judgement that a coin is biased towards heads, iff every probability in P makes that judgment, meaning that $P \subseteq \{p \in \mathcal{P}_\Omega : p(H) > 1/2\}$. More generally, an agent makes a probabilistic judgement iff their credal set P is contained in the set of probability functions corresponding to that judgement.

In this essay I will restrict myself to *regular* credal sets, that is, credal sets whose members assign some positive probability to each possibility in their domain Ω. As pointed out later on, this assumption considerably simplifies the relationship between an agent's credal set and the set of gambles they find desirable.

When $P \subseteq \mathcal{P}_\Omega$ is a credal set and $A \subseteq \Omega$ an event, I write $P(A)$ to denote the value set $\{r \in \mathbb{R} : (\exists p \in P) p(A) = r\}$, and I denote by $P(\cdot|A)$ the following conditional credal set:

$$P(\cdot|A) = \{p(\cdot|A) : p \in P\}$$

Which is defined whenever $A \neq \emptyset$, under the assumption that P is regular.

2.2 Good's theorem

We like to think evidence is valuable. For example, one should not pay to avoid free evidence. A famous theorem by Good (1967) shows that (precise) Bayesian decision theory captures these intuitions.

To make this more precise, let $\Omega = \{\omega_1, ..., \omega_n\}$ a finite possibility space.[2] Consider an agent facing a decision problem $\mathcal{A} = \{a_1, ..., a_m\}$. Let $U : \mathcal{A} \times \Omega \to \mathbb{R}$ be the agent's utility function, so the utility for option a_j when ω_i is the case is given by $U(a_j, \omega_i)$. Let $\mathcal{E} = \{E_1, ..., E_k\}$ be an arbitrary partition of events. Imagine that, at $t = 0$, the agent is offered the following choice: she can either

[2] Good's result holds for infinite possibility spaces as well.

pick some option from $\mathcal{A}$ now (at $t = 0$), or learn which $E_s \in \mathcal{E}$ is true, and then pick some option from $\mathcal{A}$ (at $t = 1$). What should she do?

Good's theorem shows that, as long as learning the events in $\mathcal{E}$ does not alter the agent's utility function, her expected utility for choosing after learning is at least as great as her expected utility for choosing without learning. Therefore, one should never pay to avoid learning free evidence.[3]

2.3 Imprecision and the value of free evidence

Trying to show that imprecise decision theory captures the same intuitions about free evidence, one is faced with a number of difficulties. First of all, while it's commonly assumed that precise Bayesian agents make choices by maximising expected utility, a number of different decision rules exist for imprecise agents (Troffaes, 2007). Furthermore, while it's straightforward to extend expected utility maximisation to sequential problems, not all IP decision rules are so easily extended. Consider the following example:[4]

Example 2.1 (Coin Toss Puzzle)**.** Jack has a coin which you know is fair. You know that Jack knows whether A is true. You know nothing about A, but judge that whether A is true is independent of the result of the coin toss. Jack paints the two sides of the coin so you can't tell which one is heads. If A is true, he writes "A" on the heads side, and "$\neg A$" on the tails side. If A is false, he writes "$\neg A$" on the heads side, and "A" on the tails side.

Let H be the event that the painted coin lands with the heads face up. Since you know the coin is fair, your starting credence in H should be $1/2$. That is, your credal set P should be such that for every $p \in P$, $p(H) = 1/2$, i.e. $P(H) = \{1/2\}$. Since you know nothing about A, you can (and perhaps should) have maximally imprecise credence in A. That is, $P(A) = (0, 1)$. Furthermore, you judge the coin toss to be independent of A. That is, if we let E_A be the event that the coin lands with the face on which "A" is painted facing up, then for every $p \in P$ it should be $p(E_A|A) = p(E_A) = 1/2$.[5]

Consider what happens after observing the painted coin toss. If the coin were to land with the "A" side up, then each $p \in P$ would take this as either evidence in favour of, or as evidence against, the fact that the coin landed heads,

[3] What makes the evidence "free" is the fact that it does not alter the agent's utility function. For an in-depth discussion of this assumption, and of Good's theorem more generally, see Kadane et al. (2008).

[4] A similar example is given by Walley (1991, pp. 298-299)

[5] Here I'm using $P(A) = (0, 1)$ instead of $P(A) = [0, 1]$ to ensure the resulting credal set is regular. This also allows me to express the judgement that the coin toss is independent of A as the fact that every $p \in P$ has $p(E_A|A) = p(E_A)$. If we had $P(A) = [0, 1]$ there would be some $p \in P$ that assigns probability 0 to A, for which the conditional probability $p(E_A|A)$ is not defined. The example can be adapted to work for any starting credal set with $P(A) = [x_1, x_2]$, where $x_1, x_2 \in (0, 1)$ and $x_1 < x_2$.

depending on $p(A)$. For each $p \in P$ we have that:

$$p(H|E_A) = \frac{p(H \cap E_A)}{p(E_A)} \tag{1}$$

$$= \frac{p(H \cap E_A|A)p(A) + p(H \cap E_A|\neg A)p(\neg A)}{p(E_A)} \tag{2}$$

$$= \frac{p(E_A|A)p(A)}{p(E_A)} = p(A) \tag{3}$$

since conditional on A, the events E_A and H are equivalent. Thus after observing E_A, your updated credal set would be maximally imprecise about H, in the sense that for every $r \in (0,1)$, there is some $p \in P(\cdot|E_A)$ with $p(H) = r$. Thus we say that your credence in H *dilates* after observing E_A.

The key feature of this example is that the two possible outcomes of the coin toss, landing with the "A" side up or with the "$\neg A$" side up, are symmetrical. Your starting credal set is also maximally imprecise about $\neg A$, meaning that for each $r \in (0,1)$ there is some $p \in P$ with $p(\neg A) = r$. And if the coin lands with the "$\neg A$" side up, we would have:

$$p(H|\neg E_A) = \frac{p(\neg E_A|\neg A)p(\neg A)}{p(\neg E_A)} = p(\neg A). \tag{4}$$

So in this case too, your credence in H will dilate.

We can use this fact to construct a sequential decision problem where the intuition that free evidence is valuable seems to fail.

Example 2.1 (continued, sequential decision problem)**.** Let a_H be the option of making a bet on the next coin toss which gains 100$ if the painted coin comes up heads, and loses 90$ otherwise. Let a_0 be the option of making no bet. Consider the following sequential decision problem: you can either learn nothing, and choose from $\mathcal{A} = \{a_0, a_H\}$ now (at $t = 0$); or you can observe the painted coin toss, see whether it lands with the "A" or "$\neg A$" face up, and then choose from $\mathcal{A} = \{a_0, a_H\}$ afterwards (at $t = 1$). The situation is represented in Figure 1.

A popular decision rule for agents with imprecise credences is Maximality.

Definition 2.1 (Maximality)**.** Let $\mathcal{A}$ be a decision problem and P a credal set. Then a_j is admissible for P from $\mathcal{A}$ iff there is no $a_i \in \mathcal{A}$ such that:

$$(\forall p \in P) EU_p(a_i) > EU_p(a_j).$$

It's not obvious how we should apply Maximality to sequential decision problems. Consider the example in Figure 1. The agent knows that at decision node 1, she would choose a_H, so at node 0 she can identify the option $\sim Learn$ with a_H. But at decision nodes 2 and 3, both a_H and a_0 are admissible to her. So at node 0, option *Learn* is not straightforwardly equivalent to one of the terminal options, and hence it's not obvious how it should be compared against $\sim Learn$.

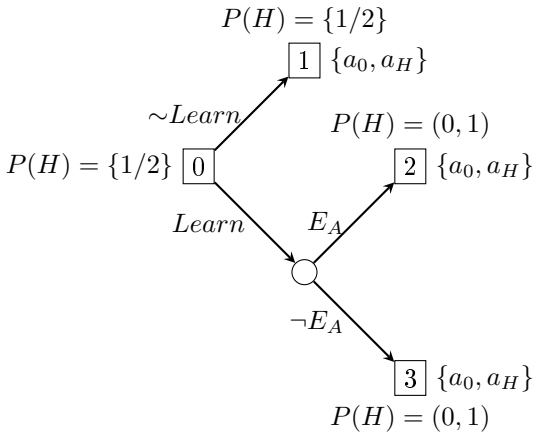

Figure 1: Representation of the Coin Puzzle as a sequential decision problem.

One way to tackle this would be to identify *Learn* with the option set $\{a_0, a_H\}$, and adapt the definition of Maximality to accommodate choices among option sets:

Definition 2.2 (Set-Maximality)**.** Let $\Theta = \{\mathcal{A}_1, ..., \mathcal{A}_n\}$ be a set of option sets and P a credal set. Then $\mathcal{A}_j$ is admissible for P from Θ iff there is no $\mathcal{A}_i \in \Theta$ such that the following two conditions hold:

1. $(\forall x \in \mathcal{A}_j)(\exists y \in \mathcal{A}_i)(\forall p \in P) EU_p(y) > EU_p(x)$,
2. $(\forall y \in \mathcal{A}_i)(\exists x \in \mathcal{A}_j)(\forall p \in P) EU_p(y) > EU_p(x)$.

Using this rule, when the agent is choosing between $\sim Learn$ and *Learn* at node 0, she is choosing between option sets $\{a_H\}$ and $\{a_H, a_0\}$. Clearly, when $P(H) = (0, 1)$, both $\{a_H, a_0\}$ and $\{a_H\}$ are admissible option sets, and thus both $\sim Learn$ and *Learn* are admissible options at node 0. The same holds if we attach a small enough price ϵ to $\sim Learn$. Hence, although paying to avoid free evidence is not mandatory, it is admissible.

There have been different proposals for extending various IP decision rules to sequential decision problems (Bradley and Steele, 2016; Kadane et al., 2008). Yet they all show that it is sometimes admissible for imprecise agents to pay to avoid learning free evidence. That is, we can find a decision problem $\mathcal{A} = \{a_1, ..., a_n\}$, a partition $\mathcal{E} = \{E_1, ..., E_k\}$, and a credal set P such that, letting *Learn* and $\sim Learn$ be sequential options defined as above, paying some small price $\epsilon > 0$ to choose $\sim Learn$ rather than *Learn* is admissible. In this sense, IP decision theories fail to capture our starting intuition that evidence is valuable.

Bradley and Steele (2016) have sought to soften the blow to IP decision theory by showing that, for a sequential version of Maximality, although learning free evidence (as opposed to paying not to learn) is not mandated, it is always

admissible. This shows a different, weaker sense in which imprecise agents value evidence: it is always admissible for them to pursue it when it's free. Hence, although imprecise agents do not value free evidence in the same way as precise ones, they still value it to some extent.

In this essay I am interested in presenting a similar response to the problem of free evidence. I want to highlight a different way in which imprecise agents value evidence, by showing that rational imprecise agents *defer* to their informed selves when it comes to their credences and betting decisions. But specifying what it means for an imprecise agent to defer to a credal set is no easy matter. Indeed, some authors have argued that imprecise probability clashes with rational deference principles, as I will discuss in the next section. So my aim in this essay will be to specify a notion of deference that does not clash with imprecise probability, and which will help us show a way in which imprecise agents value the evidence.

3 Imprecision and Deference

Our beliefs are sometimes rationally required to align with those of an expert. If your doctor believes a certain drug will treat your condition, you should also believe this, and if a trusted meteorologist predicts a hurricane is likely to hit your town tomorrow, you should also find this likely. To make this more precise, we need to specify what it means for an agent to match someone's beliefs, and also what kind of beliefs can serve as experts for a given agent. A deference principle must incorporate both these components.

It will help here to introduce some more notation. I will denote by Π the credal set of a deferring agent, writing π as shorthand to denote a singleton credal set $\Pi = \{\pi\}$. I use R as a *definite description* of the expert's credal set. This means that R may denote a different credal set R_i depending on which $\omega_i \in \Omega$ is the case (you can think of R as a function from Ω to $2^{\mathcal{P}_\Omega}$). For example, let $\Omega = \{\omega_1, \omega_2\}$, where ω_1 is the possibility that the killer was Mr. Green and ω_2 is the possibility that the killer was Mr. White. Then we could denote by R the killer's credal set, so that R_1 and R_2 would denote the credal sets of Mr. Green and Mr. White, respectively. I will write p as shorthand for the definite description of a singleton credal set $R = \{p\}$. For any random variable $X : \Omega \to \mathbb{R}$ and subset $S \subseteq \mathbb{R}$, I will write $[R(X) = S]$ for the event $\{\omega_i \in \Omega : R_i(X) = S\}$.

3.1 Precise deference principles

We are now in the position to state a classical deference principle for precise credences:

- **Reflection Principle (RP)**: Let π be an agent's precise credence function, and let p be the definite description of a precise credence function defined on the same domain. Then π defers to p iff, for every event $A \subseteq \Omega$:

$$\pi(A|\,[p(A) = s]) = s \tag{5}$$

whenever this conditional probability is defined.

The principle specifies that, conditional on the expert assigning a certain probability to an event, the agent must match them by assigning the same (conditional) probability to that event. This is in line with our intuitions about deference: you defer to someone when, conditional on them having certain credences, you also have those credences. Note that the principle also allows us to determine, for any credence function π, which credence functions it defers to. For example, Van Fraassen (1984) has argued that rational agents should defer to their future credence function.

3.2 Imprecision and deference

Imprecise analogues of the Reflection Principle have been shown to clash with the phenomenon of credal dilation, which we observed in the Coin Puzzle example. This has led some to argue against the rationality of imprecise probabilities (Topey, 2012; White, 2010).

Here is a natural generalisation of the precise Reflection Principle to the imprecise setting:

- **Value Reflection**: Let Π be an agent's credal set and R the definite description of a credal set defined on the same domain. Then Π defers to R iff for every event $A \subseteq \Omega$ and value set $S \subseteq \mathbb{R}$:

$$\Pi(A|[R(A) = S]) = S \tag{6}$$

whenever this conditional credal set is defined.

White (2010) has shown that this principle clashes with dilation. To show this, he gives the Coin Puzzle example introduced earlier, and repeated here for convenience:

Example 3.1 (Coin Toss Puzzle, Repeated)**.** Jack has a coin which you know is fair. You know that Jack knows whether A is true. You know nothing about A, but judge that whether A is true is independent of the result of the coin toss. Jack paints the two sides of the coin so you can't tell which one is heads. If A is true, he writes "A" on the heads side, and "$\neg A$" on the tails side. If A is false, he writes "$\neg A$" on the heads side, and "A" on the tails side.

Recall that in this example, you start with $\Pi(H) = \{1/2\}$ and $\Pi(A) = (0, 1)$. Furthermore, you judge the coin toss to be independent of A. So letting E_A be the event that the painted coin lands with the "A" face up, for every $p \in \Pi$ it should be $p(E_A|A) = p(E_A)$.

Instead of building a sequential decision problem from this scenario, suppose that Jack is about to toss the coin in front of you. As shown above, regardless of whether you learn E_A or $\neg E_A$, your updated credal set after observing the coin

toss will be maximally imprecise about H. This conflicts with Value Reflection. Before the coin toss, you know your credence in H will dilate, because it will do so whether you observe E_A or $\neg E_A$. Denoting your future credal set by R, this means that $[R(H) = (0, 1)]$ is just Ω. So assuming you should defer to your future credal set R, Value Reflection requires:

$$\Pi(H) = \Pi(H|\,[R(H) = (0, 1)]) = (0, 1). \tag{7}$$

This means you should not have $\Pi(H) = \{1/2\}$ before the coin toss, even though you know that the coin is fair.

To sum things up, the following four conditions are jointly inconsistent:

i Your starting credal set for the Coin Toss Puzzle has $\Pi(H) = \{1/2\}$.

ii After observing the coin toss, regardless of how it lands, your updated credal set R will have $R(H) = (0, 1)$ (Dilation).

iii You should defer to your updated credal set.

iv Π defers to R iff for every $A \subseteq \Omega$ and $S \subseteq \mathbb{R}$, $\Pi(A|\,[R(A) = S] = S)$ (Value Reflection).

Most supporters of imprecise probability agree with (i) and (ii), so they must reject either (iii) or (iv).[6] In fact, (iii) and (iv) are deeply connected. As discussed earlier, by specifying a deference principle we specify, for any given credal set, which credal sets it defers to. Therefore, whether it's true that (iii) one should defer to one's informed self (either generally or in this specific example) will depend on the deference principle (iv) we endorse. Note also that claim (iii), that one should defer to one's informed self, is what I presented in the previous section as a way to show that evidence is valuable for imprecise agents. Hence in this essay I will reject (iv) Value Reflection, and my aim will be to specify, and justify as best I can, an IP deference principle that is consistent with (i) - (iii).

4 Justifying a Deference Principle

Before presenting a deference principle for agents with imprecise credences, I should say a bit more about how such a principle can be justified. One way to do so is to list desiderata which the principle should satisfy, and then check whether and to what extent a candidate principle satisfies them.

The first desideratum is that the principle should capture some intuitions about what it means to defer to someone. This follows from the discussion in Section 2: showing that an imprecise agent's credal state is related to the credal state of her informed self as specified by a deference principle is our way to show that imprecise agents value evidence.

[6] Although see (Bradley and Steele, 2014) for a discussion of update rules which avoid dilation.

- **(D1) Captures deference intuitions**: the principle should capture some of our intuitions about what it means to defer to someone. For example, it's natural to think that, if one defers to an expert, then conditional on the expert having a certain opinion one should also have the same opinion.

The second desideratum is also related to our discussion in Section 2: the candidate deference principle should establish that imprecise agents defer to their updated selves, thus showing that evidence is valuable for imprecise agents. At first pass, we can formulate this using the same setup as Good's theorem.

- **(D2) Defer to informed self**: Let Π be a regular credal set defined over Ω, $\mathcal{E} = \{E_1, ..., E_k\}$ be a partition of events, and denote by R the credal set obtained by updating Π on whichever $E_s \in \mathcal{E}$ is true. Then Π should defer to R.

Note that satisfying (D2) would show that the principle can accommodate the Coin Toss Puzzle, unlike Value Reflection. Since the dilated credal set is obtained by updating the initial credal set, if (D2) is satisfied, then the initial credal set will defer to the updated one.

The third desideratum is that the principle follows from other, independently motivated assumptions. For example, precise probabilistic coherence has been justified as a requirement of rationality by showing that an agent who violates it is vulnerable to a Dutch book (de Finetti, 1964; Pettigrew, 2020). Since we have independent reasons to believe that being Dutch-bookable is a fault of rationality, this gives us reason to think coherence is a norm of rationality.

- **(D3) Independent motivation**: The principle should follow from independently motivated assumptions.

The fourth and final desideratum requires that, when only precise credences are involved, our imprecise deference principle should collapse to a reasonable precise deference principle. Often the precise case is easier to study, and if we have good reasons to reject a deference principle in the precise case, these may overpower the reasons we have for supporting it in its more general, imprecise form.

- **(D4) Non-revisionist**: The principle should collapse to a reasonable precise deference principle when both the agent and the expert credal sets are singletons.

5 Two IP Deference Principles

I will define imprecise deference principles in terms of gambles which an agent finds desirable. So it will be useful to introduce some notation to talk about these gambles. A gamble is a function $X : \Omega \to \mathbb{R}$, where $X(\omega_i)$ denotes the value paid by the gamble when ω_i is the case. Given an option a_j and a utility

function U, we can define a corresponding gamble $X_j = U(a_j, \cdot)$ whose payout at ω_i is just the utility of option a_j if ω_i is the case.

When $X, Y : \Omega \to \mathbb{R}$ are gambles, I write $X + Y$ to denote the following gamble on Ω:

$$(X + Y)(\omega) = X(\omega) + X(\omega) \tag{8}$$

and define $X - Y$ similarly. If $A \subseteq \Omega$ is an event and $X : \Omega \to \mathbb{R}$ is a gamble, I write AX to denote the following gamble:

$$AX = \begin{cases} X(\omega) & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

I also use the symbols $\geq, =$ to denote the following binary relationships between gambles:

$$X \geq Y \iff \text{for every } \omega \in \Omega, X(\omega) \geq Y(\omega) \tag{10}$$

$$X = Y \iff \text{for every } \omega \in \Omega, X(\omega) = Y(\omega) \tag{11}$$

and write $X \neq Y$ whenever it's not the case that $X = Y$. I denote by $\mathcal{L}(\Omega)$ the set of all gambles on Ω.

If P is an agent's credal set, denote by $D_P \subseteq \mathcal{L}(\Omega)$ the agent's *set of (strictly) desirable gambles*, defined by:

$$D_P = \{X : \text{ for every } p \in P, \operatorname{Exp}_p(X) > 0\} \tag{12}$$

Intuitively, these are just the gambles that a rational agent with credal set P would be disposed to accept. Every set of desirable gambles generated in this way from a regular credal set respects the following *coherence constraints*:

$$0 \notin D \tag{C1}$$

$$X \geq 0, X \neq 0 \implies X \in D \tag{C2}$$

$$X \in D, \lambda > 0 \implies \lambda X \in D \tag{C3}$$

$$X, Y \in D \implies (X + Y) \in D \tag{C4}$$

A set of desirable gambles that respects the above constraints is said to be *coherent*.[7]

We are finally ready to state the two IP deference principles I want to put forward:[8]

[7] If P is not regular, then defining D_P as above does not guarantee that the coherence axiom (C2) is respected. A common way to address this is to define $D_P = \{X : \text{ for every } p \in P, p(X) > 0\} \cup \{X \in \mathcal{L}(\Omega) : X \geq 0, X \neq 0\}$. However, this definition complicates the relationship between the set $D_{P(\cdot|A)}$ of desirable gambles according to the conditional credal set $P(\cdot|A)$, and the set $\{X : XA \in D_P\}$ of gambles which are desirable for P conditional on A. This relationship is central to the results presented in this essay. I will leave the problem of extending these results to accommodate nopn-regular credal sets to another time.

[8] The form and name of these principles were inspired by Dorst et al. (2021). Later we will see that, when all probabilities involved are precise, both STT and WTT are equivalent to the principle given by Dorst et al., called "Total Trust".

- **Strong Total Trust (STT)**: Let Π be a regular credal set, and R be the definite description of a credal set defined on the same domain. Π defers to R iff for every gamble $X : \Omega \to \mathbb{R}$, we have:

$$X \in D_{\Pi(\cdot|[X \in D_R])} \tag{13}$$

  whenever this conditional credal set is defined, and where $[X \in D_R] = \{\omega_i \in \Omega : X \in D_{R_i}\}$. If this is the case, we say that Π *S-Trusts* R.

- **Weak Total Trust (WTT)**: Let Π be a regular credal set, and R be the definite description of a credal set defined on the same domain. Π defers to R iff for every gamble $X : \Omega \to \mathbb{R}$:

$$-X \notin D_{\Pi(\cdot|[X \in D_R])} \tag{14}$$

  whenever this conditional credal set is defined. If this is the case, we say that Π *W-Trusts* R.

Note that, if Π S-Trusts R, then Π also W-Trusts R. This is because, if $X \in D$, then $-X \notin D$, whenever D is coherent. Let's see how these principles fare against the desiderata (D1-D4) listed above.

5.1 D1: Capture deference intuitions

Starting from (D1), the intuition behind both STT and WTT is similar to that behind Value Reflection. In both cases, conditional on the expert's imprecise credence having a certain property, the agent's imprecise credence should match it in some way. In the case of Value Reflection the property in question is the set of prevision values assigned to a random variable, whereas in the case of STT it is the disposition to accept a gamble corresponding to that variable. So we can give the following informal definition of STT: an agent defers to an expert when, conditional on the expert finding a gamble desirable, the agent finds that gamble desirable. WTT requires the agent to match the expert in a weaker sense: an agent defers to an expert when, conditional on the expert finding a gamble desirable, the agent does not find it desirable to sell that gamble.

Both STT and WTT have an interesting alternative formulation. Before introducing it, it is useful to define the notion of strict preference between options.

Definition 5.1 (Strict preference)**.** Let P be a credal set, U be a utility function, and a_1, a_2 be two options. We say that P *strictly prefers* a_1 to a_2 under U, when:

$$(X_1 - X_2) \in D_P \tag{15}$$

where $X_1 = U(a_1, \cdot)$ is the gamble corresponding to option a_1, and $X_2 = U(a_2, \cdot)$ is the gamble corresponding to option a_2.

Let Π be an agent's credal set, and U the agent's utility function, and let R be the definite description of another credal set. Assume as usual that P is regular.

Consider a binary decision problem $\mathcal{A} = \{a_1, a_2\}$, and assume that learning R's preferences (under U) on $\mathcal{A}$ does not affect the agent's utility function on $\mathcal{A}$. Define a new option s_1 which is equal to a_2 whenever R strictly prefers a_2 to a_1 under U, and equal to a_1 otherwise. Define s_2 analogously.

$$s_1 = \begin{cases} a_2 & \text{if } R \text{ strictly prefers } a_2 \text{ to } a_1 \text{ under } U \\ a_1 & \text{otherwise.} \end{cases} \tag{16}$$

$$s_2 = \begin{cases} a_1 & \text{if } R \text{ strictly prefers } a_1 \text{ to } a_2 \text{ under } U \\ a_2 & \text{otherwise.} \end{cases} \tag{17}$$

You can think of s_1 as a "black box" option which, if R has a definite preference in $\mathcal{A}$, contains R's preferred option, and otherwise contains a_1. Denote by $[s_1 \neq a_1]$ the event $\{\omega_i \in \Omega : R_i \text{ strictly prefers } a_2 \text{ to } a_1\}$. Denote by $[s_1 \neq a_1]$ the event $\{\omega_i \in \Omega : R_i \text{ strictly prefers } a_2 \text{ to } a_1\}$. We can characterise STT in terms of these black box options as follows:

Proposition 1. *Π S-Trusts R iff for every binary decision problem $\mathcal{A} = \{a_1, a_2\}$, the following hold:*

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

Proof. Assuming the utility of the options in $\mathcal{A}$ is not affected by learning facts about the expert's preferences, we can treat options as gambles. That is, to each option a_j corresponds a gamble X_j which pays $U(a_j, w_i)$ when w_i is the case. Similarly, write S_j for the gamble which pays $U(s_j, w_i)$ when w_i is the case, where s_j is defined as above. Then we have:

$$\begin{aligned} &S_1 - X_1 = \\ &= X_1 \left[(X_2 - X_1) \notin D_R\right] + X_2 \left[(X_2 - X_1) \in D_R\right] - X_1 \\ &= X_2 \left[(X_2 - X_1) \in D_R\right] - X_1 \left[(X_2 - X_1) \in D_R\right] \\ &= (X_2 - X_1) \left[(X_2 - X_1) \in D_R\right] \end{aligned}$$

We say that Π strictly prefers s_1 to a_1 iff $(S_1 - X_1) \in D_\Pi$, and by the above this is the case iff:

$$(X_2 - X_1) \left[(X_2 - X_1) \in D_R\right] \in D_\Pi \tag{18}$$

$$\iff \pi((X_2 - X_1)[(X_2 - X_1) \in D_R]) > 0 \text{ for all } \pi \in \Pi \tag{19}$$

If $[(X_2 - X_1) \in D_R] \neq \emptyset$, is equivalent to:

$$\pi((X_2 - X_1)|[(X_2 - X_1) \in D_R]) > 0 \text{ for all } \pi \in \Pi \tag{20}$$

$$\iff \pi(X_2 - X_1) > 0 \text{ for all } \pi \in \Pi(\cdot|[(X_2 - X_1) \in D_R]) \tag{21}$$

where these conditional probabilities are all defined because of the assumption that Π is regular. Clearly Π S-Trusts R iff condition (21) holds for every pair of gambles such that $[(X_2 - X_1) \in D_R] \neq \emptyset$. And since by definition $[s_1 \neq a_1] = [(X_2 - X_1) \in D_R]$, this gives the result. □

A similar characterisation can be given for W-Trust:

Proposition 2. Π *W-Trusts* R *iff for every binary problem* $\mathcal{A} = \{a_1, a_2\}$*, the following hold:*

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

Proof. Analogous to Proposition 1. □

This gives us another intuitive way to think about deference. Π defers to R when it values R's preferences. For STT, this means that a black box containing R's preferred option when R has a definite preference, and containing a_j otherwise, is at least as good as a_j according to Π. For WTT, it means that this black box is not definitely worse than a_j according to Π.

5.2 D2: Defer to informed self

To capture the intuition that evidence is valuable, we want to show that a rational agent should defer to their updated credences, both in the case of STT and in the case of WTT. Since STT is the stronger constraint, is suffices to show that agents S-Trust their updated credences.

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*.*

Proof. Assume by way of contradiction that Π does not S-Trust R. Then there is some gamble X such that $X \notin D_{\Pi(\cdot|[X \in D_R])}$, where $[X \in D_R] \neq \emptyset$. Under the assumption that Π is regular, this is equivalent to:

$$\pi(X) \leq 0 \text{ for some } \pi \in \Pi(\cdot|[X \in D_R]) \tag{22}$$
$$\iff \pi(X|[X \in D_R]) \leq 0 \text{ for some } \pi \in \Pi \tag{23}$$
$$\iff \pi(X[X \in D_R]) \leq 0 \text{ for some } \pi \in \Pi \tag{24}$$
$$\iff X\,[X \in D_R] \notin D_\Pi \tag{25}$$

We know R is obtained by updating Π on whichever $E_s \in \mathcal{E}$ is true, so we can rewrite this as:

$$X \bigcup_{s:X \in D_{\Pi(\cdot|E_s)}} E_s \notin D_\Pi \tag{26}$$

And because the members of $\mathcal{E}$ are mutually exclusive, this is the same as:

$$\sum_{s:X \in D_{\Pi(\cdot|E_s)}} XE_s \notin D_\Pi \tag{27}$$

So there must be some E_s such that $XE_s \notin D_\Pi$, while also $X \in D_{\Pi(\cdot|E_s)}$. But as above, $X \in D_{\Pi(\cdot|E_s)}$ is equivalent to $XE_s \in D_\Pi$, contradiction. □

The fact that imprecise agents defer to their informed selves shows some sense in which they value evidence. This holds regardless of the fact that Good's theorem seems to fail for these agents when we extend imprecise decision theory to sequential problems. In fact, note how the above result does not require us to settle on a specific decision rule for imprecise agents, nor does it require to exend this rule to the sequential case. All that is needed to define our deference principles and to prove Proposition 3 is the notion of desirability of gambles, which is fairly uncontroversial.[9]

5.3 D3: Independent motivation

The precise restriction of STT/WTT (discussed in more detail below) has been justified by appealing to diachronic Dutch book considerations by Dorst et al. (2021). The argument shows that, if π does not defer to p as described by the precise restriction of STT/WTT, then we can find two decision problems $\mathcal{X}, \mathcal{X}'$, both containing a zero option, such that adding up π's choice from $\mathcal{X}$ and p's choice from $\mathcal{X}'$ we obtain a gamble that is negative on every $\omega_i \in \Omega$. Conversely, if π defers to p, no such pair of decision problems exists (Dorst et al., 2021).

Note that this Dutch book argument seems to only motivate the deference principle under the assumption that p is the agent's future credence. If that's the case, a diachronic Dutch book of the kind described above does indeed look like a rationality fault: if I am offered each decision problem at the appropriate time, I will incur a sure loss. However, the argument does not motivate the principle as a general deference constraint. For example, I may defer to my doctor's opinions on the outcome of a certain treatment. But the fact that you can offer me and my doctor two decision problems on this domain, and that our combined choices ensure a loss, is not obviously a sign of irrationality on my part. Hence, it's not clear that my credences should relate to my doctor's credences as prescribed by Dorst's deference principle.

Here I will also restrict myself to cases where R is the agent's future credal set. An interesting subclass is the one we focused on in our discussion of the value of free evidence. There, we assumed that the possible values of R are obtained by updating Π on the elements of a partition $\mathcal{E}$. In these cases, we already know by Propositions 3 and 4 that coherence requires Π to S-Trust R. But there are cases in which R is the agent's future credal set, and yet its values are not obtained by updating Π on the elements of a partition. Here is an example involving precise credences:

Example 5.1. Two coins are about to be tossed. The first coin is fair. If the first coin comes up tails, the second toss will also be fair, whereas if the

[9] This notion of desirability arguably does impose some constraints on the decision rule. For example, an imprecise agent with credal set Π who uses the Γ-maximin decision rule may choose the constant gamble 0 over some other gamble X, even though $X \in D_\Pi$. Hence supporters of Γ-maximin may find our notion of desirability inadequate. But there are independent reasons to reject Γ-maximin (Seidenfeld, 2004), and the two most popular IP decision rules, E-admissibility and Maximality, coincide on binary decision problems, and are consistent with our notion of desirability (Troffaes, 2007).

first coin comes up heads, the second toss will be rigged to ensure heads comes up. Therefore the possibility space is $\Omega = \{H_1H_2, T_1H_2, T_1T_2\}$. Let π be your starting credence function over Ω, and let p denote your credence function before observing the first toss that comes up heads, if at least one toss does come up heads; and let p denote your credence function after observing both tosses otherwise. That is:

$$p = \begin{cases} p_1 = \pi & \text{if} \quad H_1H_2 \\ p_2 = \pi(\cdot|T_1) & \text{if} \quad T_1H_2 \\ p_3 = \pi(\cdot|T_1T_2) & \text{if} \quad T_1T_2 \end{cases} \tag{28}$$

The key feature of this example is that, if $\{T_1H_2\}$ is the case, the agent's future credence p_2 does not have the information that $[p = p_2]$ as part of its total evidence. To see this, note that $[p = p_2] \equiv \{T_1H_2\}$, and p_2 only has T_1 as evidence. Indeed, the fact that p_2 is not certain that $[p = p_2]$ is reflected by the probability assignment $p_2(\{T_1H_2\}) = 1/2$.[10]

Should one defer to one's future credence in these cases? On the face of it, p has at least as much evidence as π does, and possibly has more. So we would expect an agent who values evidence to defer to p. If that's right, this example shows a way in which precise agents value the evidence that is not captured by Good's theorem. I will end this section by discussing how this issue is relevant for our choice of IP deference principle.

5.4 D4: Non-revisionist

If both Π and R are singleton sets containing a single regular probability function (call their elements π and p), then both STT and WTT are equivalent the following deference principle, given by Dorst et al. (2021).

- **Total Trust** π defers to p iff for every gamble X:

$$\pi(X|\,[p(X) \geq 0]) \geq 0 \tag{29}$$

  whenever this conditional prevision is defined. If this is the case, we say that π *Totally Trusts* p.

The reason Dorst et al. (2021) introduce Total Trust is that Precise Reflection is known to be problematic in cases where the expert credence is *modest*, i.e. when for some possible expert credence p_i, $p_i([p = p_i]) < 1$. This is the case in Example 5.1, where Precise Reflection imposes the following constraint:

$$\begin{aligned} &\pi([p(T_1H_2) = p_2(T_1H_2)]|\,[p(T_1H_2) = p_2(T_1H_2)]) \\ &= p_2([p(T_1H_2) = p_2(T_1H_2)]) \quad \text{(by Reflection)} \\ &= p_2(T_1H_2) = 1/2 < 1 \end{aligned}$$

[10] For an overview of this phenomenon, and a number of similar examples, see (Schervish et al., 2004).

which violates the ratio formula. On the other hand, it can be shown that π does Totally Trust (and therefore also S/W-Trust) p in this example. So if we think that, in order to value the evidence, agents should defer to their future credences in cases like Example 5.1, this shows that Total Trust is better than Reflection as a precise deference principle.

The same considerations may help us rule out some IP deference principles. Here is an alternative principle given in the literature:

- **Identity Reflection**[11]: Let Π be an agetnt's credal set and R the definite description of a credal set defined on the same domain. Then Π defers to R iff:

$$\Pi(\cdot\,|\,[R = R_i]) = R_i \tag{30}$$

  whenever this conditional credal set is defined. If this is the case, we say Π *I-Reflects* R.

It's easy to show that the following analogue of Proposition 3 holds for Identity Reflection:

Proposition 4. *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 I-Reflects R.*

So under Identity Reflection, we also have that imprecise agents defer to their informed selves under the assumptions of Good's theorem. We can also show Identity Reflection is no weaker than STT and WTT.

Proposition 5. *Let Π be a regular credal set, and R the definite description of a credal set defined on the same domain. If Π I-Reflects R, then Π S-Trusts R.*

Proof. Assume Π I-Reflects R. Then let $X : \Omega \to \mathbb{R}$ such that $[X \in D_R] \neq \emptyset$. Then we have:

$$X[X \in D_R] = X \bigcup_{R_i : X \in D_{R_i}} [R = R_i] \tag{31}$$

$$= \sum_{R_i : X \in D_{R_i}} X[R = R_i]. \tag{32}$$

But for every R_i such that $X \in D_{R_i}$, we know by Identity Reflection that $\Pi(\cdot|[R = R_i]) = R_i$, and therefore $X \in D_{\Pi(\cdot|[R=R_i])}$. Since Π is regular, this in turn implies $X[R = R_i] \in D_\Pi$. So each gamble in the sum above belongs to D_Π, and by coherence their sum belongs to D_Π. Thus we have:

$$X[X \in D_R] \in D_\Pi \tag{33}$$

which, since $[X \in D_R] \neq \emptyset$ and Π is regular, implies $X \in D_{\Pi(\cdot|[X \in D_R])}$. □

[11] Similar principles are mentioned in (Topey, 2012; Schoenfield, 2012). The name and formulation given here is due to Moss (2021).

As a consequence of this, one direction of Proposition 1 goes through for Identity Reflection:

Proposition 6. *If* Π *I-Reflects* R*, then for every binary problem* $\mathcal{A} = \{a_1, a_2\}$*, the following hold:*

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

where s_1, s_2 *are the black box options defined in* (16)*,* (17)*.*

However, Identity Reflection is strictly stronger than STT. That is, we can find Π and R such that Π S-Trusts R, but Π does not I-Reflect R. To see this, note that in the setup of Example 5.1, Identity Reflection fails in the same way as Precise Reflection does, since conditioning on $[p(T_1H_2) = p_2(T_1H_2)]$ is the same as conditioning on $[p = p_2]$. Hence, in Example 5.1, π does not I-Reflect p. However, π does S-Trust p in that example. If we think agents should defer to their informed selves in cases like Example 5.1, this gives us a reason to prefer STT/WTT over Identity Reflection as a deference principle for imprecise credences.

6 Looking ahead

The material presented above has a number of limitations. This section gives a quick overview of each limitation, and discusses prospects for overcoming them.

6.1 Beyond regularity

The assumption that credal sets are regular is quite restrictive. However, this assumption greatly simplifies the relationship between conditional credal sets and their set of conditional desirable gambles. If P is regular, then by defining $D_P = \{X : p(X) > 0 \text{ for all } p \in P\}$ we get:

$$D_{P(\cdot|A)} = \{X : AX \in D_P\} \tag{34}$$

whenever $A \neq \emptyset$. That is, for every event A on which we might be interested in conditioning P, a gamble X is desirable for $P(\cdot|A)$ iff the conditional gamble XA, which pays X when A occurs and is called off otherwise, is desirable for P.

If P is not regular, then defining D_P as above does not ensure that this set is coherent. In particular, coherence axiom (C2) may be violated, since it's possible that for some gamble X, $X \geq 0$ and $X \neq 0$, but $p(X) = 0$ for some $p \in P$, meaning that $X \notin D_P$. [12] This is is usually addressed by defining the set of desirable gambles corresponding to a credal set P as follows:

$$D_P^+ = D_P \cup \{X : X \geq 0 \text{ and } X \neq 0\} \tag{35}$$

[12] Note that because of this, neither set of desirable gambles in (34) is coherent on Ω, since $P(\cdot|A)$ is not regular on Ω. But both sets are coherent when restricted to A.

which clearly ensures (C2) is respected. However, this complicates the relationship between $D^+_{P(\cdot|A)}$ and $\{X : AX \in D^+_P\}$. For one, note that even if $A \neq \emptyset$, it could be that $0 \in P(A)$. If that's the case, then our definition of conditional credal set, $P(\cdot|A) = \{p(\cdot|A) : p \in P\}$, will leave the conditional credal set undefined. Consequently, $D^+_{P(\cdot|A)}$ will also be undefined. But the set $\{X : AX \in D^+_P\}$ is well-defined when $0 \in P(A)$: it's just the set $\{X : AX \geq 0 \text{ and } AX \neq 0\}$.

The definition of conditional credal set is commonly extended to accommodate cases where $0 \in P(A)$, by defining:

$$P(\cdot|A) = \begin{cases} \{p(\cdot|A) : p \in P\} & \text{if } 0 \notin P(A) \\ \{p : p \in \mathcal{P}_\Omega \text{ and } p|_A \in \mathcal{P}_A\} & \text{if } 0 \in P(A), A \neq \emptyset \end{cases} \tag{36}$$

leaving $P(\cdot|A)$ undefined just in case $A = \emptyset$. But even though this solves the domain problem above, we still can't recover the nice equality (34). Consider what happens when $0 \in P(A)$, and A is such that $AX \geq 0, AX \neq 0$, even though $X(\omega) < 0$ for some $\omega \notin A$. Then we have that $AX \in D^+_P$, by coherence axiom (C2), even though $X \notin D^+_{P(\cdot|A)}$, because $P(\cdot|A)$ is the set of all probability functions on Ω assigning probability 1 to A, and hence $D^+_{P(\cdot|A)} = \{X : X \geq 0 \text{ and } X \neq 0\}$. So we have the unpleasant result that the conditional gamble XA is desirable for the credal set P, even though the unconditional gamble X is not desirable for the conditional credal set $P(\cdot|A)$.

This issue highlights a difference of informativeness between sets of conditional desirable gambles, and conditional credal sets. The set $\{X : AX \in D_P\}$ of gambles desirable for P conditional on A is obtained by *ruling out* the possibilities outside of A as irrelevant to judgements of desirability. That is, two gambles X and Y such that $X(\omega) = Y(\omega)$ for every $\omega \in A$ are indistinguishable in terms of their desirability conditional on A. Meanwhile the set $D_{P(\cdot|A)}$ is obtained by *assigning probability 0* to the possibilities outside of A, which does not rule them out as possibilities. Hence a gamble X that pays 0 on every possible world, and a gamble Y which pays 0 on A but is positive for some $\omega \notin A$, will be treated differently by $D_{P(\cdot|A)}$: the former is not desirable, whereas the latter is.

Assigning probability 0 to a possibility ω and ruling ω out as not possibile are two different attitudes. The regularity assumption essentially collapses the two, and thus solves the problems above. But I think a better solution would be to introduce this distinction in the credal set perspective. If we don't want to conflate impossibility with having probability 0, it's natural to read off impossibility from the domain of the functions in the credal set. That is, if a credal set contains functions defined on Ω, then the elements of Ω are possibilities, regardless of their probability. Then to be in line with conditional desirability of gambles, if P is a credal set on Ω and $A \subseteq \Omega$, we can let $P(\cdot|A)$ be a set of probabilities defined on A, rather than on Ω:

$$P(\cdot|A) = \begin{cases} \{p(\cdot|A)|_A : p \in P\} & \text{if } 0 \notin P(A) \\ \{p : p \in \mathcal{P}_A\} & \text{if } 0 \in P(A), A \neq \emptyset \end{cases} \tag{37}$$

If P is defined on a domain A and $A \subseteq B$, we can still define the set of gambles on B that are desirable for P in a relatively harmless way by:

$$D_P^B = \{X \in \mathcal{L}(B) : p(X|_A) > 0 \text{ for all } p \in P\} \cup \{X \in \mathcal{L}(B) : AX \geq 0, AX \neq 0\}$$

This set won't be coherent on B, but it will be coherent on the credal set's domain A. And it's easy to show that, using these definitions, we recover the desired equality. For every credal set P defined over Ω, and every $A \subseteq \Omega$ such that $A \neq \emptyset$, we have:

$$D^{\Omega}_{P(\cdot|A)} = \{X \in \mathcal{L}(\Omega) : AX \in D_P^{\Omega}\} \tag{38}$$

At the moment this is the solution I prefer, and variants of the results given in this essay go through under this definition of conditional credal set. But I need to think a bit more about the philosophical implications of thinking about conditionalisation in this way.

6.2 Beyond binary choice

An open question is whether STT/WTT can be modified to produce interesting constraints, of the kind expressed by Propositions 1 and 2, for arbitrary decision problems, instead of being limited to cases where the option set is binary. Analogous results are given for arbitrary decision problems in the precise case by Dorst et al. (2021).

I suspect that this limitation to binary decision problems has to do with the fact that STT and WTT are defined in terms of desirable gambles. That is, whether an agent with credal set P S-Trusts or W-Trusts R depends only on the corresponding sets of desirable gambles D_P and D_R. But different credal sets may produce the same sets of desirable gambles, even when they make importantly different probabilistic judgements, as shown by the following example.

Example 6.1. Let $\Omega = \{\omega_1, \omega_2\}$, and let $p_1, p_2 : \Omega \to \mathbb{R}$ the probability functions with $p_1(\omega_1) = .9 = p_2(\omega_2)$. Let $P = \{p_1, p_2\}$ and $P' = \{p \in \mathcal{P}_\Omega : 0.1 \leq p(\omega_1) \leq 0.9\}$. We have that $D_P = D_{P'}$, even though the two credal sets are importantly different. The first credal set makes the probabilistic judgement that the probability of ω_1 is either in $[.9, 1]$ or in $[0, .1]$ (because every probability in the set agrees with this judgement) whereas the second credal set does not.

To capture the full expressive power of credal sets, we can think of desirability in terms of *choiceworthiness*, along the lines of much recent imprecise probability literature (Van Camp, 2018; De Bock and De Cooman, 2019). As with desirability, there are some slight inconveniences when defining choiceworthiness for an arbitrary credal set, so I will restrict myself to regular credal sets here for simplicity. If P is a regular credal set and $\mathcal{X} \subseteq \mathcal{L}(\Omega)$ a set of gambles, we can define the set $C_P(\mathcal{X})$ of choiceworthy options from $\mathcal{X}$ according to P:

$$C_P(\mathcal{X}) = \{X \in \mathcal{X} : \text{ for some } p \in P, X \text{ maximises } p \text{ in } \mathcal{X}\} \tag{39}$$

The fact that a gamble X is desirable for P can then be expressed as the fact that, in the set of gambles $\{X, 0\}$, the set of choiceworthy options for P is $C_P(\{X, 0\}) = \{X\}$. More generally, from the set of desirable gambles D_P associated to P we can determine, for every binary set of gambles $\mathcal{X} = \{X_1, X_2\}$ the set of choiceworthy options $C_P(\mathcal{X})$:

$$C_P(\{X_1, X_2\}) = \begin{cases} \{X_1\} & \text{if } (X_1 - X_2) \in D_P, \\ \{X_2\} & \text{if } (X_2 - X_1) \in D_P, \\ \{X_1, X_2\} & \text{otherwise} \end{cases} \tag{40}$$

However, D_P does not suffice to determine the set of choiceworthy options $C_P(\mathcal{X})$ for an arbitrary n-ary $\mathcal{X}$, as shown by the following example.

Example 6.2. Let Ω, P, and P' be defined as in Example 6.1. Let $\mathcal{X} = \{X_1, X_2, X_3\}$ a set of gambles on Ω such that $X_1(\omega_1) = 1$, $X_1(\omega_2) = -0.1$; $X_2 = 0.9 - X_1$; $X_3(\omega_1) = 1$, $X_3(\omega_2) = -1$. Then $C_P(\mathcal{X}) = \{X_1, X_2\}$ and $C_{P'}(\mathcal{X}) = \{X_1, X_2, X_3\}$, even though $D_P = D_{P'}$.

A special case is the case of singleton credal sets, where we have that $D_p = D_{p'}$ iff $p = p'$ iff $C_p = C_{p'}$, which explains why binary choice is equivalent to n-ary choice for precise probabilities. If we want to work with arbitrary (regular) credal sets, however, it makes sense to define deference principles in terms of choiceworthiness rather than desirability. I am currently working on developing such principles.

6.3 Independent justification

One of the main shortcomings of the deference principles presented in this essay is that they are not independently motivated in general. In particular cases, such as when the expert is the agent's future self, under the assumptions of Good's theorem, STT and WTT are motivated by coherence. However, I have shown cases involving precise probabilities where one should intuitively defer to one's future self, even though the assumptions of Good's theorem don't hold.

Dorst et al. (2021) give a Dutch Book argument for their deference principle which covers such precise cases, but it would be nice to give an argument for STT/WTT that applies more generally to imprecise cases. I have made very little progress on this for the moment. It's also worth noting that Dorst et al. (2021) give a second, accuracy-based characterisation of their deference norm. This is not immediately available in the imprecise case, since there are currently no strictly proper scoring rules for imprecise probabilities.

A related open question is whether one really ought to defer to their future self under some minimal assumptions about one's utility function and update rules, or whether there are more significant constraints for when our future selves should be regarded as epistemic authorities.

7 Conclusion

Rational agents value the evidence. This intuition can be captured by Good's theorem, which shows rational agents never pay to avoid free evidence. But Good's theorem seems to fail for imprecise probabilities, raising a worry that imprecise probabilities are inadequate for a theory of rationality.

This essay argued that our intuitions about the value of evidence are best captured by appealing to deference principles. I defined two deference principles for imprecise probabilities, STT and WTT, and showed that, under the assumptions of Good's theorem, imprecise agents defer to their informed selves. This shows a way in which they value evidence. Furthermore, there are cases involving precise probabilities in which Good's theorem does not apply, but where our intuitions about the value of evidence suggest that a rational agent should defer to their future self. This further supports the conclusion that the value of evidence is best analysed in terms of deference.

I have shown STT/WTT follow from coherence under the assumptions of Good's theorem. But it remains an open question whether there is a way to justify STT/WTT when these assumptions don't hold, for example, by giving a Dutch Book argument of the kind given by Dorst et al. (2021). Another open question is whether STT/WTT can be modified to produce interesting constraints, of the kind expressed by Propositions 1 and 2, for arbitrary decision problems, instead of being limited to cases where the option set is binary. Finally, whether Π defers to R under STT and WTT depends only on the corresponding sets of desirable gambles D_Π and D_R. But different credal sets may produce the same set of desirable gambles even when they make importantly different probabilistic judgements. A future goal would be to define an IP deference principle that is sensitive to these differences.

References

Bradley, S. and Steele, K. (2014). Uncertainty, learning, and the "problem" of dilation. *Erkenntnis*, 79:1287–1303.

Bradley, S. and Steele, K. (2016). Can free evidence be bad? value of information for the imprecise probabilist. *Philosophy of Science*, 83(1):1–28.

De Bock, J. and De Cooman, G. (2019). Interpreting, axiomatising and representing coherent choice functions in terms of desirability. In *International Symposium on Imprecise Probabilities: Theories and Applications*, pages 125–134. PMLR.

de Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In Kyburg, H. and Smokler, H., editors, *Studies in Subjective Probability*, pages 53–118. Wiley New York. Translated by H. Kyburg (original published in 1937).

Dorst, K. (2020). Evidence: A guide for the uncertain. *Philosophy and Phenomenological Research*, 100(3):586–632.

Dorst, K., Levinstein, B. A., Salow, B., Husic, B. E., and Fitelson, B. (2021). Deference done better. *Philosophical Perspectives*, 35(1):99–150.

Good, I. J. (1967). On the principle of total evidence. *The British Journal for the Philosophy of Science*, 17(4):319–321.

Kadane, J. B., Schervish, M., and Seidenfeld, T. (2008). Is ignorance bliss? *The Journal of Philosophy*, 105(1):5–36.

Moss, S. (2021). Global constraints on imprecise credences: Solving reflection violations, belief inertia, and other puzzles. *Philosophy and Phenomenological Research*, 103(3):620–638.

Pettigrew, R. (2020). *Dutch book arguments*. Cambridge University Press.

Schervish, M. J., Seidenfeld, T., and Kadane, J. B. (2004). Stopping to reflect. *The Journal of Philosophy*, 101(6):315–322.

Schoenfield, M. (2012). Chilling out on epistemic rationality: A defense of imprecise credences (and other imprecise doxastic attitudes). *Philosophical Studies*, 158(2):197–219.

Seidenfeld, T. (2004). A contrast between two decision rules for use with (convex) sets of probabilities: γ-maximin versus e-admissibility. *Synthese*, 140(1/2):69–88.

Topey, B. (2012). Coin flips, credences and the reflection principle. *Analysis*, 72(3):478–488.

Troffaes, M. C. (2007). Decision making under uncertainty using imprecise probabilities. *International journal of approximate reasoning*, 45(1):17–29.

Van Camp, A. (2018). Choice functions as a tool to model uncertainty (phd dissertation). *Phd Dissertation*.

Van Fraassen, B. C. (1984). Belief and the will. *The Journal of Philosophy*, 81(5):235–256.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*, volume 42. Springer.

White, R. (2010). Evidential symmetry and mushy credence. *Oxford studies in epistemology*, 3:161–186.