Trusting and Valuing Infinite Frames

Brian Weatherson

2024

Abstract: I show that some recent results concerning finite probability frames do not generalise to infinite frames.

Say a probability frame is an ordered pair ⟨W, P⟩ such that W is a set (intuitively, of worlds), and P is a function from W to probability functions defined on w. Say that π is a probability function defined over w. Intuitively, π is Novice's credence function, and P maps members of W to the credence a special person called The Expert has at each world. The philosophical question in the background, and it will stay in the background for this note, is what relationship must hold between π and ⟨W, P⟩ to say that Novice defers to The Expert.

Kevin Dorst (2019) made some suggestions about what this relationship must look like, but it turned out these weren't quite right. In Dorst et al. (2021), it is shown that if W is finite, then the following two claims are equivalent. (I'll state the claims formally, then explain my notation.)

Total Trust $E(X \mid \{w: E(X, P(w)) \geqslant t\}, \pi) \geqslant t$

Value If O is a set of options, s is a recommended strategy for O, and o is a member of O, then $E(s, \pi) \geqslant E(o, \pi)$.

There is a bit there to unpack. E is the expectation function, so E(X, Pr) is the expectation of X according to Pr. We'll generalise this a little to say that E(X | p, Pr) is the expectation of X according to Pr(• | p). So here's what Total Trust says. Take any random variable X. (In this context, a random variable is just a function from W to $\mathbb{R}$.) Update π on the proposition that consists of all and only worlds where The Expert at that world has an expected value for X at least equal to t. After that update, Novice also expects X's value to be at least t.

A decision problem is just a set of random variables. Intuitively, the problem is that Novice has to choose which bet to take from the set O, with the return being the value of the chosen bet at the actual world. A strategy for Novice is to defer the decision to The Expert. Formally, a strategy s is a function from W to O; it picks a random variable, i.e., a bet, at each world. For a strategy to be *recommended*, it has to satisfy two constraints. First, if $P(i) = P(j)$, then $s(i) = s(j)$. A strategy can't be more discriminating than The Expert's credences. For the second constraint, associate each strategy s with a random variable S such that $S(w) = s(w)(w)$. That is, the value of S at w is the return at w of the member of O that s selects at w. The constraint then says that for any $O_1 \in O$, $E(s(w), P(w)) \geqslant E(O_1, P(w))$. That is, at w, The Expert does not think there is a member of O that in expectation does better than $s(w)$. Given all that, Value says that The Novice does not expect any member of O to do better than any recommended strategy.

One of the things Dorst et al. (2021) show is that if W and O are finite, then Total Trust and Value are equivalent. As they put it, π Totally Trusts a frame $\langle W, P\rangle$ iff it Values that frame. What I'll show is that this equivalence breaks down when we drop the finiteness assumptions. Indeed, it breaks down even when W is countably infinite.

Start with a frame I'll call **Coin**. A fair coin will be flipped repeatedly until it lands Tails. Let F be a random variable such that $F = x$ iff the coin is flipped x times. (If the coin never lands Tails, we'll stipulate that $F = 1$. Since this has probability 0, it doesn't make a difference to what follows, but this case will matter below.) Novice knows these facts about F, so $\pi(F = x) = 2^{-x}$. If $F = x$, then The Expert knows $F \geqslant x$, and updates on that. That is, $P(F = x) = \pi(\bullet|F \geqslant x)$. For any positive integer i, let O_i be the random variable that takes value 0 at $F = j$ when $j \leqslant i$, and value 2^i at $F = j$ when $j > i$. Let O be the set of each O_i. The strategy s such that $s(F = i) = O_i$ is recommended, as can be easily checked. But $E(s) = 0$, while for any $o \in O$, $E(o, \pi) = ½$. So Value fails on **Coin**.

On the other hand, π does Totally Trust **Coin**. [Note to self, I need a proof here.]

So π Totally Trusts **Coin**, but doesn't Value it. So the equivalence between Total Trust and Value fails here. But you might very reasonably object on two scores. First, the value function used to generate the counterexample was unbounded, and we know that unbounded value functions lead to all sorts of paradoxes. Second, I didn't just make W infinite, I made O infinite as well, so this isn't a minimal generalisation of the original claim. It turns out that if we put both these constraints on, then the equivalence fails in the other direction: It is possible to get a frame that π Values, but does not Totally Trust.

Call the following frame **Bentham**. Again, a coin will be flipped until it lands Tails. If it ever lands

Dorst, Kevin. 2019. "Evidence: A Guide for the Uncertain." *Philosophy and Phenomenological Research* 100 (3): 586–632. doi: 10.1111/phpr.12561.

Dorst, Kevin, Benjamin A. Levinstein, Bernhard Salow, Brooke E. Husic, and Branden Fitelson. 2021. "Deference Done Better." *Philosophical Perspectives* 35 (1): 99–150. doi: 10.1111/phpe.12156.

In progress.