The Principle of Indifference and the Principal Principle are Incompatible

J. Dmitri Gallow

The Principle of Indifference (POI) says that, in the absence of evidence, you should distribute your credences evenly. The Principal Principle (PP) says that, in the absence of evidence, you should align your credences with the chances. PETTIGREW (2016) appears to accept both the PP and the POI. Many other authors write as though Bayesians are free to accept both of these principles. HAWTHORNE ET AL. (2017) even go so far as to argue that the PP *implies* the POI.[1] PETTIGREW (2018) and TITELBAUM and HART (2018) found the flaws in their argument, but they left untouched the Bayesian who accepts both the POI and the PP. This Bayesian has contradicted themselves, since the POI and the PP are incompatible. Abiding the POI means violating the PP. So Bayesians cannot accept both principles; they must choose which, if either, to endorse.

1 The Principle of Indifference

Let Ω be the set of possibilities over which your credences are defined.[2] If there are only finitely many possibilities in Ω, then the POI says that each possibility should be given the same credence. That is, if C is a rational initial, or *ur-prior*, credence function—a credence function which it would be rational to hold in the absence of evidence—then, for each $\omega \in \Omega$, $C(\omega) = 1/\#\Omega$.[3]

I'll suppose that the possibilities over which your credences are defined can be generated from some underlying language, which, for the

Draft of May 1, 2020; Word Count: 3,903

1. More carefully, they argue that the PP implies that your credence in every atomic proposition should be 1/2, which is strictly weaker than the POI.

2. More carefully, your credences are defined over *subsets* of Ω. I'll suppose throughout that you have a credence in every singleton $\{\omega\}$.

3. Notation: I use '$\#S$' for the cardinality of the set S.

sake of simplicity, I'll take to be a truth-functional propositional language. If the *atomic* propositions in this language are finite in number, $A_1, A_2, \dots, A_N$, then the set of possibilities, Ω, will be finite. For we can associate each possibility $\omega \in \Omega$ with a unique *state description*. A *state description* is a conjunction of the form $\pm A_1 \wedge \pm A_2 \wedge \cdots \wedge \pm A_N$, where each $\pm A_i$ is either the atomic proposition A_i or its negation. Since the language is truth-functional, a state description settles the truth-value of every other proposition in the language, so that any two possibilities which agree about a state description agree *tout court*. So there's no need to have multiple possibilities in which the same state description is true; and we can take the possibilities and the state descriptions to correspond one-to-one. Then, if there are N atomic sentences in your language, there will be 2^N state descriptions, and the cardinality of Ω will be 2^N. Then, the POI will say that your credence in each singleton $\{\omega\} \subset \Omega$ should be $1/2^N$.

If there are countably many possibilities, $\#\Omega = \aleph_0$, the POI conflicts with countable additivity and normalization (two standard axioms of probability theory). Countable additivity says that, for any sequence of pairwise disjoint subsets of Ω, $P_1, P_2, P_3, \dots$, your credence in the union of the P_i's should be equal to the sum of your credence in each P_i. And normalization says that your credence in Ω itself must be 100%. To see that POI violates either countable additivity or normalization, consider the singleton of every possibility in Ω: $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}, \dots$. POI says that your credence in each of these singletons must be the same. It can either be positive or zero. If positive, $C(\{\omega_i\}) = \alpha > 0$, then $\sum_{i=1}^{\infty} C(\{\omega_i\}) = \sum_{i=1}^{\infty} \alpha = \infty$. But the union of all the singletons $\{\omega_i\}$ is just Ω itself. So countable additivity require would require $C(\Omega) = \infty$, in violation of normalization. On the other hand, if you give credence zero to each $\omega_i \in \Omega$, then countable additivity will require your credence in Ω to be $\sum_{i=1}^{\infty} C(\{\omega_i\}) = \sum_{i=1}^{\infty} 0 = 0$. But normalization requires that $C(\Omega) = 1$. So either way, POI will lead you to violate either countable additivity or normalization.

Defenders of the POI could give up countable additivity, retreating to the strictly weaker *finite* additivity, which only applies to *finite* sequences of pairwise disjoint subsets of Ω.[4] Or they could instead allow that, in some cases, distributing your credences *perfectly evently* will vio-

4. That is: finite additivity says that, if $P_1, P_2, \dots, P_N$ are pairwise disjoint subsets of Ω, then $C\left(\bigcup_{i=1}^{N} P_i\right) = \sum_{i=1}^{N} C(P_i)$.

late the probability axioms, but still insist that rationality requires you to distribute your credences *sufficiently evenly*, while still complying with countable additivity.[5]

With uncountably many possibilities in Ω, it is not entirely clear what it even *means* to spread your credences among them equally. Every uncountable subset of Ω is as large as every other, so we might try to say that every uncountable subset gets the same credence. But this would lead us back into violations of additivity or normalization. Take some uncountably infinite subsets of Ω, P and Q, such that P and Q are disjoint and their union is Ω itself, $P \cup Q = \Omega$. Then, additivity (countable or finite) would tell us that $C(\Omega) = C(P) + C(Q)$, and POI (so understood) would tell us that $C(\Omega) = C(P) = C(Q)$. But this implies that $C(\Omega) = 0$, in violation of normalization.

So, in the absence of some further structure on the set Ω, it's not even clear how to interpret the POI. The usual way the POI is implemented when the set of possibilities is uncountably infinite involves imposing some additional structure on Ω by finding some way of parameterizing the possibilities in Ω. That is, we find some random variable, V which maps every possibility in $\omega \in \Omega$ to some real number, $V(\omega) \in \mathbb{R}$.Then, we can assign to each value v in the range of the variable V a *credence density*, $\rho_V(v)$. This density function *doesn't* say what your credence that $V = v$ is.[6] If you abide by the POI, your credence that V takes on any particular value, v, will have to be zero. Instead, $\rho_V(v)$ says how *dense* your credence is at $V = v$. Think about it like this: for any narrow interval $[v, v + \epsilon]$, the ratio $C(V \in [v, v + \epsilon])/\epsilon$ is the density of your credence over the interval $[v, v + \epsilon]$. By taking the limit of this ratio as ϵ goes to zero, we get the density of your credence at the *point* $V = v$, $\rho_V(v)$.

With a credence density function, ρ_V, we can get your credence that V takes on any particular values by integrating over ρ_V. For instance, your credence that V is between a and b will be given by $\int_a^b \rho_V(v)\, dv$. And, in general, for any measurable set of values $\mathbf{v}$, your credence that

5. This is the approach adopted by Williamson (2010). He says that the Shannon entropy of your credences should be *sufficiently* high—while allowing that it may not be maximal, because it could be that *no* probability function has maximal entropy.

6. I use '$V = v$' to stand for the set of possibilities which V maps to v, $V = v \overset{\text{def}}{=} \{\omega \in \Omega \mid V(\omega) = v\}$. Likewise, '$V \in \mathbf{v} \overset{\text{def}}{=} \{\omega \in \Omega \mid V(\omega) \in \mathbf{v}\}$.

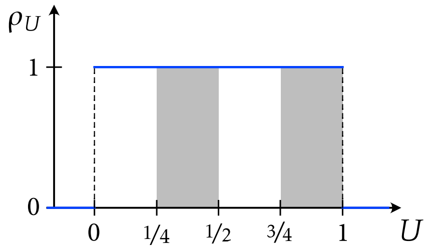

FIGURE 1: The uniform credence density over U. Your credence that U lies in the set $\mathbf{u} = [1/4, 1/2] \cup [3/4, 1]$ is given by the integral $\int_{\mathbf{u}} \rho_U(u)\ \mathrm{d}u$, which is the area under the curve $\rho_U(u)$ shown in grey.

V is within $\mathbf{v}$ is given by $\int_{\mathbf{v}} \rho_V(v)\ \mathrm{d}v$.[7] Then, the POI is implemented by saying that your credences should have a *uniform density*. That is: every value of v should have exactly the same credence *density*.

For instance: we can characterize each possibility by what percentage of space is unoccupied in that possibility. Call that variable U (for *u*noccupied). The variable U can take on values between 0 and 1. Then, POI says that the density of your credence should be *uniform* over these values. This uniform credence density is shown in figure 1.

2 The Principal Principle

David LEWIS (1980)'s *principal principle* says something about the conditional credences of a rational initial, or *ur-prior*, credence function, C. In particular, it says: if P is any proposition, t is some future time, $Ch_t(P) = x$ is the proposition that the time t chance of P is x, for some real number $x \in [0, 1]$, and E is any time t admissible proposition, then

$$C(P \mid Ch_t(P) = x \wedge E) = x$$

The time t won't be important in my discussion, so I'll fix t to be some future time and omit explicit mention of t in the remainder. Likewise, the admissible proposition E won't play any important role. So I'll assume only that the trivial proposition Ω is admissible at t, and set $E = \Omega$. This

7. In general, we could characterize the possibilities in Ω with any finite number of real-valued variables, $V_1, V_2, \ldots, V_N$. Then, instead of having a density function on $\mathbb{R}$, we'd have a density function on $\mathbb{R}^N$. In the interests of simplicity, I'll focus on the simplest case, where $N = 1$.

leads to the following principle, which I'll call 'the PP' from here on out:

$$C(P \mid Ch(P) = x) = x \tag{PP}$$

The PP governs your *conditional* credences; but I'll suppose that these conditional credences place a constraint on your unconditional credences, *via* the product rule, which says that, for any propositions P and Q, your credence in $P \wedge Q$ is equal to the product of your credence that P *given* Q and your credence that Q.

$$C(P \wedge Q) = C(P \mid Q) \cdot C(Q)$$

Then, so long as $C(Q) > 0$, your conditional credence in P, given Q, is the ratio of your unconditional credence in $P \wedge Q$ and your unconditional credence in Q, $C(P \mid Q) = C(P \wedge Q)/C(Q)$. So a constraint on your conditional credences will have consequences for your unconditional credences.

Rational credences are probabilities, so, conditional on any proposition whatsoever, if your credence in P is x, then your credence in $\neg P$ must be $1 - x$. So if, conditional on $Ch(P) = x$, your credence in P is x, then, conditional on $Ch(P) = x$, your credence in $\neg P$ must be $1 - x$. So, supposing that $C(Ch(P) = x) > 0$,

$$\frac{C(P \mid Ch(P) = x)}{C(\neg P \mid Ch(P) = x)} = \frac{x}{1-x}$$

$$\frac{C(P \wedge Ch(P) = x)/C(Ch(P) = x)}{C(\neg P \wedge Ch(P) = x)/C(Ch(P) = x)} = \frac{x}{1-x}$$

$$\frac{C(P \wedge Ch(P) = x)}{C(\neg P \wedge Ch(P) = x)} = \frac{x}{1-x}$$

$$C(P \wedge Ch(P) = x) = \frac{x}{1-x} \cdot C(\neg P \wedge Ch(P) = x) \tag{1}$$

Equation 1 follows from the PP. It will be important in §3 below.

Some Humeans do not accept the PP because it conflicts with their metaphysical commitments.[8] Nonetheless, those Humeans are happy to accept a principle they call 'the *new* principle'. Where the original principal principle implores you to defer to the chances, the new principle implores you to defer to the chances *conditional* on the true the-

8. See Hall (1994), Lewis (1994), and Thau (1994).

ory of chance. The differences between the new principle and the original principal principle won't be relevant to anything I say here. If you favor the new principle, you can simply interpret '$Ch(P) = x$' as the proposition that the time t chance of P—*conditional on the true theory of chance*—is x.[9] So understood, it will follow from the new principle that $C(P \wedge Ch(P) = x) = (x/(1-x)) \cdot C(\neg P \wedge Ch(P) = x)$.

Suppose you want your credences to be defined over uncountably many propositions of the form $Ch(A) = x$—one for each of the uncountably many real numbers, x, between 0 and 1. Then, so long as your credences are real valued, you'll have to assign a credence of zero to uncountably many of the propositions $Ch(A) = x$. If your credence in $Ch(A) = x$ is zero, then the product rule will not impose any constraint on the relationship between $C(A \mid Ch(A) = x)$ and $C(A \wedge Ch(A) = x)$. Lewis was not concerned with this, because he allowed rational credences to take on infinitesimal values.[10] So he thought that, even when you're spreading your credences over uncountably many propositions, you needn't give a credence of zero to any of them. If we agree with him about this, then perhaps the PP is already general enough. But I've been persuaded that Lewis was wrong to rely upon infinitesimals.[11] If, like me, you want your credences to be real-valued, then you should be looking for a natural generalization of the PP for the case where you have credences over uncountably many chance propositions.

Even though your credence that the chance of P is x will be zero for any particular choice of x, your credence that the chance of P lies within an *interval* of values $[x, x+\epsilon]$ (with $\epsilon > 0$) can be non-zero, no matter how small the interval $[x, x+\epsilon]$. So a natural generalization of the PP says that a rational ur-prior credence in P, given that the chance of P lies in some interval $[x, x+\epsilon]$, is within the interval $[x, x+\epsilon]$:

(PP*) $$x \leqslant C(P \mid Ch(P) \in [x, x+\epsilon]) \leqslant x+\epsilon$$

If your credence in P, given $Ch(P) \in [x, x+\epsilon]$, is in the interval $[x, x+\epsilon]$, then your credence in $\neg P$, given $Ch(P) \in [x, x+\epsilon]$, is within

9. *Cf.* Hall and Arntzenius (2003).

10. See Lewis (1980, pp. 267–8).

11. See Williamson (2007), Easwaran (2014), and Hájek (ms, §7).

the interval $[1-x-\epsilon, 1-x]$:

$$1-x-\epsilon \leqslant C(\neg P \mid Ch(P) \in [x, x+\epsilon]) \leqslant 1-x$$

Following the same steps from our derivation of equation 1 above, we get that, for any positive ϵ, no matter how small,

$$\begin{aligned} C(P \wedge Ch(P) \in [x, x+\epsilon]) &\leqslant \frac{x+\epsilon}{1-x-\epsilon} \cdot C(\neg P \wedge Ch(P) \in [x, x+\epsilon]) \\ \text{and } C(P \wedge Ch(P) \in [x, x+\epsilon]) &\geqslant \frac{x}{1-x} \cdot C(\neg P \wedge Ch(P) \in [x, x+\epsilon]) \end{aligned}$$

Divide both sides of these inequalities by ϵ, and take the limit as ϵ goes to zero. Thereby, we get that the *density* of your credence in the conjunction $P \wedge Ch(P) = x$ must be $x/(1-x)$ times the *density* of your credence in the conjunction $\neg P \wedge Ch(P) = x$,

$$\rho(P \wedge Ch(P) = x) = \frac{x}{1-x} \cdot \rho(\neg P \wedge Ch(P) = x) \tag{2}$$

Equation 2 follows from the PP*. It will be important in §3 below.

3 The Incompatibility

Suppose that your credences are defined over a language which includes only two atomic propositions: A and $Ch(A) = 1/3$. Then, there will be four possible state descriptions: $A \wedge Ch(A) = 1/3$, $A \wedge Ch(A) \neq 1/3$, $\neg A \wedge Ch(A) = 1/3$, and $\neg A \wedge Ch(A) \neq 1/3$. The POI requires that your credence in $A \wedge Ch(A) = 1/3$ is the same as your credence in $\neg A \wedge Ch(A) = 1/3$. It says that both should be equal to 1/4th:

$$C(A \wedge Ch(A) = 1/3) = C(\neg A \wedge Ch(A) = 1/3) = 1/4 \tag{3}$$

Whereas the PP implies that your credence in $A \wedge Ch(A) = 1/3$ must be one half of your credence in $\neg A \wedge Ch(A) = 1/3$ (this follows from equation 1):

$$\begin{aligned} C(A \wedge Ch(A) = 1/3) &= \left(\frac{1/3}{1-1/3}\right) \cdot C(\neg A \wedge Ch(A) = 1/3) \\ &= (1/2) \cdot C(\neg A \wedge Ch(A) = 1/3) \end{aligned} \tag{4}$$

The POI requires (3), while the PP requires (4). It's not possible to satisfy both (3) and (4). So the POI and the PP are incompatible.

Perhaps the incompatability only arises because we chose an impoverished language. So let's enrich our language by allowing in any finite number of *non-chancy* atomic propositions, $A_1, A_2, \ldots A_N$, along with any finite number of *chancy* atomic propositions of the form $Ch(A_i) = x$, where A_i is an atomic proposition and x is a real number between 0 and 1. With multiple of these chance propositions included, a defender of the POI will have to change the way they understand a *state description*. To see why, notice that if we say that a state description is a conjunction whose conjuncts include every atomic proposition or its negation, then there could be state descriptions including the conjuncts $Ch(A_i) = x$ and $Ch(A_i) = y$, for $x \neq y$. This conjunction is known to be impossible, so it should receive credence zero, but since the POI requires you to give every state description the same credence, the POI would require your credence in this known impossibility to be positive.

The solution is to say that, for each *non-chancy* atomic proposition, A_i, a state description includes either A_i or its negation as a conjunct, and, for each atomic proposition A_j which has a chance,[12] the state description includes exactly one *chancy* atomic proposition of the form $Ch(A_j) = x$, or else it includes the the negation of every proposition of that form, $\bigwedge_x Ch(A_j) \neq x$.

Now, suppose that there's at least one atomic proposition A such that $Ch(A) = x$ is a chancy atomic proposition with $x \neq 1/2$. Let $\mathcal{S}_{A,x}$ be the set of all state descriptions which include both A and $Ch(A) = x$. Then, by finite additivity, your credence in the conjunction $A \wedge Ch(A) = x$ must be equal to the sum of your credence in every state description in $\mathcal{S}_{A,x}$.

$$C(A \wedge Ch(A) = x) = \sum_{S \in \mathcal{S}_{A,x}} C(S)$$

If you satisfy the POI, then every state description will receive the same credence, so each summand in the above sum will be the same—call it 'α'. So your credence that $A \wedge Ch(A) = x$ will be α times the number of state descriptions in $\mathcal{S}_{A,x}$.

(5) $$C(A \wedge Ch(A) = x) = \alpha \cdot \#\mathcal{S}_{A,x}$$

12. I say that the atomic proposition A_j *has a chance* iff there is some chancy atomic proposition of the form $Ch(A_j) = x$.

In the same way, let $\mathcal{S}_{\neg A,x}$ be the set of all state descriptions which include both $\neg A$ and $Ch(A) = x$. Again, by finite additivity, your credence in the conjunction $\neg A \wedge Ch(A) = x$ must be equal to the sum of your credence in every state description in $\mathcal{S}_{\neg A,x}$. By the POI, each of these state descriptions must be the same value, α. So your credence that $\neg A \wedge Ch(A) = x$ will be α times the number of state descriptions in $\mathcal{S}_{\neg A,x}$.

(6) $$C(\neg A \wedge Ch(A) = x) = \alpha \cdot \#\mathcal{S}_{\neg A,x}$$

But the number of state descriptions in $\mathcal{S}_{A,x}$ must be equal to the number of state descriptions in $\mathcal{S}_{\neg A,x}$. Take any $S \in \mathcal{S}_{A,x}$, replace 'A' with '$\neg A$', and you have a state description $S^* \in \mathcal{S}_{\neg A,x}$. This associates each $S \in \mathcal{S}_{A,x}$ with a unique $S^* \in \mathcal{S}_{\neg A,x}$, so $\#\mathcal{S}_{A,x} \leqslant \#\mathcal{S}_{\neg A,x}$. For every $S^* \in \mathcal{S}_{\neg A,x}$, replace '$\neg A$' with '$A$', and you have a state description $S \in \mathcal{S}_{A,x}$. This associates each $S^* \in \mathcal{S}_{\neg A,x}$ with a unique $S \in \mathcal{S}_{A,x}$, so $\#\mathcal{S}_{\neg A,x} \leqslant \#\mathcal{S}_{A,x}$. So $\#\mathcal{S}_{A,x} = \#\mathcal{S}_{\neg A,x}$. And this, together with (5) and (6), implies that

(7) $$C(A \wedge Ch(A) = x) = C(\neg A \wedge Ch(A) = x)$$

But, since we stipulated that $x \neq 1/2$, (7) is incompatible with the PP, which requires $C(A \wedge Ch(A) = x) = (x/(1-x)) \cdot C(\neg A \wedge Ch(A) = x)$ (recall equation 1).

But perhaps the incompatibility only arises because we have no more than a *finite* number of chancy atomic propositions of the form $Ch(A) = x$. So let's open the door to uncountably many possibilities. For each $x \in [0, 1]$, we will have a possibility in which the chance of A is x. Then, for each atomic proposition A which has a chance, we may parameterize this uncountably infinite space with a variable, V, defined to be $\mathbf{1}_{\neg A} + Ch(A)$. Here, '$\mathbf{1}_{\neg A}$' is the truth-value of $\neg A$, and '$Ch(A)$' is the chance of A. So, if V is between 0 and 1, then A is true and the value of V is the chance of A. And, if V is between 1 and 2, then A is false, and the value of V is the chance of A plus 1.[13] The POI tells you to have a uniform credence density over the values of V, as in figure 2.

13. We can have a similar variable for every atomic proposition which has a chance. By assuming independence amongst all of the atomic propositions (an assumption the defender of the POI should be happy to grant), we can generate a joint credence density function over all of these variables from the uniform marginal densities. But we can bring out the incompatibility of the POI and the PP* by just looking at the marginal distribution over V.

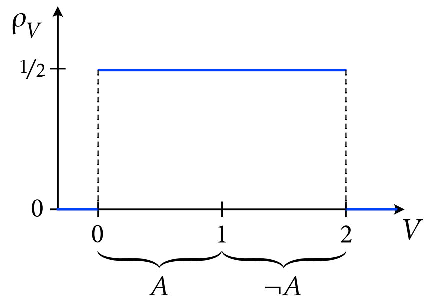

FIGURE 2: The uniform density over $V = \mathbf{1}_{\neg A} + Ch(A)$ is required by the POI, but is incompatible with the PP*.

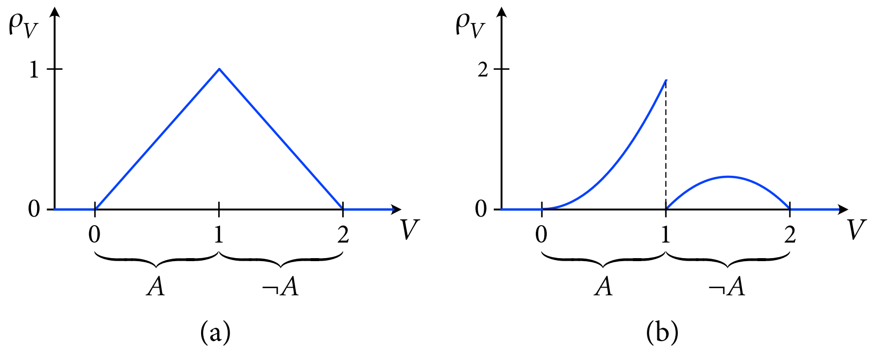

FIGURE 3: Two sample credence densities over $V = \mathbf{1}_{\neg A} + Ch(A)$ which abide the PP. (In figure 3a, ρ_V is v between 0 and 1, $2 - v$ between 1 and 2, and 0 elsewhere. In figure 3b, ρ_V is $2v^2$ between 0 and 1, $6v - 2v^2 - 4$ between 1 and 2, and 0 elsewhere.)

But this is incompatible with the PP* (the generalization of the PP for situations in which the number of potential chance hypotheses is uncountably infinite). For the PP* requires that, for any v between 0 and 1,

(8) $$\rho_V(v) = \frac{v}{1-v} \cdot \rho_V(1+v)$$

(Equation 8 follows from from equation 2, which itself follows from the PP*, as we saw in §2.) But the uniform credence density shown in figure 2 sets $\rho_V(v) = \rho_V(1+v) = 1/2$ for every value of v between 0 and 1. So the uniform credence density will violate equation 8 for every value of v other than $v = 1/2$. So the uniform credence density does not abide the PP*. (I've shown two sample credence densities which abide the PP* in figure 3.)

4 Further Discussion

Following Lewis, I have formulated the PP and the PP* as principles which, just like the POI, constrain rational *initial* credences. These principles have nothing to do with how your credences are disposed to change upon receiving evidence. Nonetheless, some defenders of the POI may see hidden in the PP and the PP* a vestige of the principle of conditionalization, according to which you should be disposed to update your credences by conditioning on any newly acquired evidence. Those defenders of the POI may wish to reject the PP and the PP* as I've explicitly formulated them, but accept nearby principles which, instead of constraining your initial *conditional* credences, constrain the credences you are *disposed* to adopt, upon learning what the chance of a proposition is. For instance, in response to an unrelated puzzle, Wallmann and Williamson (2020, p. 3) suggest that the principal principle should be understood as saying that $C_{Ch(P)=x}(P) = x$, where $C_{Ch(P)=x}$ is the credence function you are disposed to adopt, upon learning that $Ch(P) = x$ and no more. The principle of conditionalization says that $C_{Ch(P)=x}$ should be C conditional on $Ch(P) = x$, so this proposed principle agrees with the PP when conjoined with conditionalization. Williamson, however, rejects conditionalization. Instead, he says that, when your evidence imposes constraints on your credences, you should adopt a probability which meets those constraints and which otherwise distributes its probability evenly (or, perhaps, *sufficiently* evenly). Following Jaynes (1957),Williamson calls this updating norm *the principle of maximum entropy*. (The name comes from the fact that the evenness of your credence can be measured by its entropy.)

Notice that this response does not call into question my main contention here, which is that the PP and the POI are incompatible. The response acknowledges the incompatibility, but rejects the PP. It attempts to mitigate this rejection by showing that there is some *other* norm governing the connection between your credence and chance which is compatible with the POI.

In many cases, the principle of maximum entropy will agree with the principle of conditionalization. If C is defined over E, and you receive evidence which imposes the constraint that E receive credence 1, then the updating norm of maximum entropy will require that C_E is C

conditioned on E.[14] That is: if you already have a credence in E, and your evidence imposes the constraint that your credence in E be 1, then the norm of maximum entropy will tell you to update by conditioning on the proposition E. So, if learning that $Ch(P) = x$ imposes the constraint that $Ch(P) = x$ be assigned a credence of 1, then $C_{Ch(P)=x}(P)$ will be equal to $C(P \mid Ch(P) = x)$. In that case, since the POI won't set $C(P \mid Ch(P) = x)$ equal to x unless $x = 1/2$, the principle of maximum entropy won't set $C_{Ch(P)=x}(P)$ equal to x unless $x = 1/2$, either.

Williamson doesn't actually say that, upon learning that E, you should update by imposing the sole evidential constraint that E be assigned credence 1. That's how things work for the *non-chancy* propositions. But there's something special about learning chance propositions. Learning a chance proposition like $Ch(P) = x$ doesn't only impose the evidential constraint that $C(Ch(P) = x) = 1$. It additional imposes the evidential constraint that your credence in P be equal to x.[15] Then, the proposed revision of the PP is trivially satisfied: for any proposition P and any real number x, $C_{Ch(P)=x}(P)$ will be x, by stipulation.

What if you just learn that the chance of P lies within some *range* of values? In that case, Williamson says that your credence that P must lie within that range.[16] For illustration, suppose that you begin with the credence distribution shown in figure 2, and you learn that $Ch(A) \geqslant 1/2$. In that case, your credence that A is currently $1/2$, so you currently satisfy the constraint to have a credence in the interval $[1/2, 1]$. Moreover, your credence that A is independent of your credence that $Ch(A) \geqslant 1/2$, so $C(A \mid Ch(A) \geqslant 1/2) = 1/2$, and you will still satisfy the constraint to have a credence in the interval $[1/2, 1]$ after conditioning on the chance proposition $Ch(A) = 1/2$. So Williamson won't advise you to increase your credence in A at all, even though your expectation of the chance of A has risen from $1/2$ to $3/4$ths.

We could try to get around this problem by insisting that the evidence $Ch(A) \geqslant 1/2$ imposes the constraint that your credence that A

14. See Seidenfeld (1986)'s 'Result$_1$', on page 471.

15. "Learning $Ch(P) = x$ does not merely impose the constraint $C(Ch(P) = x) = 1$, but also the constraint $C(P) = x$" (Williamson, 2010, p. 79, with minor notational changes).

16. More generally, he says that your credence that P must lie within the convex hull of the numbers which might, for all your evidence has to say, be the chance of P. See §3.3.1 of Williamson (2010).

equal 3/4ths. But then, if you were to go on to learn $Ch(A) \leq 3/4$, we would presumably want to impose the new constraint that your credence in A be your (new) expectation of the chance of A, $C(A) = 5/8$. At that point, the constraints on your credences would be inconsistent. It's clear that the constraint $C(A) = 3/4$ should be ditched and that the constraint $C(A) = 5/8$ should take its place, though it's less clear whether there's any principled story to be told about why. In any case, this kind of approach seems to me to confuse *evidence*—which is the *input* to an updating rule—with the *rational response to* that evidence—which is the *output* of an updating rule. To see why, think about what would happen, had you first learnt that $Ch(A) \leqslant 3/4$, and then learn that $Ch(A) \geqslant 1/2$. In that case, we would have to say that the evidence $Ch(A) \leqslant 3/4$ imposes the constraint that $C(A) = 3/8$, and that the evidence $Ch(A) \geqslant 1/2$ imposes the constraint that $C(A) = 5/8$. But why should changing the order in which you receive the evidence about chance make a difference to the constraints which that evidence imposes? Whether you already know that $Ch(A) \geqslant 1/2$ should make a difference to which credences you adopt when you learn that $Ch(A) \leqslant 3/4$; but I don't see why the evidence you've already received would make any difference to the constraint which the piece of evidence $Ch(A) \leqslant 3/4$ *itself* imposes on your credences. We should be able to specify the evidential constraints imposed by a new piece of evidence in a way which is independent of your prior credences and your pre-existing evidence.

Whether we adopt this proposal or not, if you update your credences in the way Williamson advises, you will violate the rule of conditionalization whenever you stand to learn something about the chances. So, whenever you stand to learn something about the chances, you will be susceptible to a Dutch book strategy, as Teller (1973, 1976) and Lewis (1999) have shown. Williamson recognizes this, but contends that susceptibility to a Dutch book strategy is no vice. He argues for this as follows: suppose you are about to listen to the defense. You know that they will only present evidence which supports the defendant's innocence. So, if you know you're rational, then you know that your credence in the defenant's innocence will go up, no matter what you hear. So, after their defense, you'll sell back a bet on the defendant's guilt for less than you paid for it.[17] But Williamson is wrong that rationality will compel you to lower your credence in the defendant's guilt no matter what you

17. See (Williamson, 2010, §4.4).

hear. An exceptionally weak defense should make you more confident that the defendant is guilty (think: *that's the best they could do?*).

In sum: the POI and the PP are incompatible. The defender of the POI can offer surrogate chance deference principles besides the PP. However, these surrogates are not very plausible. They will tell you to not change your credence that A, even when your expectation of the chance of A has been raised. We could try to get around this problem, but only by customizing the evidential constraints on a case-by-case basis. And, even if we did this, the surrogate would expose you to diachronic exploitability.

References

Easwaran, Kenny, 2014. Regularity and Hyperreal Credences. *The Philosophical Review*, 123(1):1–41. [6]

Hájek, Alan, ms. Staying Regular? [6]

Hall, Ned, 1994. Correcting the Guide to Objective Chance. *Mind*, 103(412):505–517. [5]

Hall, Ned and Arntzenius, Frank, 2003. On What We Know About Chance. *The British Journal for the Philosophy of Science*, 54(2):171–179. [6]

Hawthorne, James, Landes, Jürgen, and Williamson, Jon, 2017. The Principal Principle Implies the Principle of Indifference. *The British Journal for the Philosophy of Science*, 68(1):123–131. [1]

Jaynes, E. T., 1957. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630. [11]

Lewis, David K., 1980. A Subjectivist's Guide to Objective Chance. In Richard C. Jeffrey, editor, *Studies in Inductive Logic and Probability*, volume II, pages 263–293. University of California Press, Berkeley. [4], [6], [11]

Lewis, David K., 1994. Humean Supervenience Debugged. *Mind*, 103(412):473–490. [5]

Lewis, David K., 1999. Why Conditionalize? In *Papers in Metaphysics and Epistemology*, volume 2, chapter 23, pages 403–407. Cambridge University Press, Cambridge. [13]

Pettigrew, Richard, 2016. *Accuracy and the Laws of Credence*. Oxford University Press, Oxford. [1]

Pettigrew, Richard, 2018. The Principal Principle Does Not Imply the Principle of Indifference. *The British Journal for the Philosophy of Science*. [1]

Seidenfeld, Teddy, 1986. Entropy and Uncertainty. *Philosophy of Science*, 53(4):467–491. [12]

Teller, Paul, 1973. Conditionalization and Observation. *Synthese*, 26(2):218–258. [13]

Teller, Paul, 1976. Conditionalization, observation, and change of preference. In W. L. Harper and C. A. Hooker, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume I, pages 205–253. D. Reidel Publishing Company, Dordrecht. [13]

Thau, Michael, 1994. Undermining and Admissibility. *Mind*, 103(412):491–503. [5]

Titelbaum, Michael G and Hart, Casey, 2018. The Principal Principle Does Not Imply the Principle of Indifference, Because Conditioning on Biconditionals Is Counterintuitive. *The British Journal for the Philosophy of Science*. ISSN 0007-0882. doi:10.1093/bjps/axy011. [1]

Wallmann, Christian and Williamson, Jon, 2020. The Principal Principle and Subjective Bayesianism. *European Journal for Philosophy of Science*, 10(3):1–14. [11]

Williamson, Jon, 2010. *In Defense of Objective Bayesianism*. Oxford University Press, Oxford. [3], [11], [12], [13]

Williamson, Timothy, 2007. How probable is an infinite sequence of heads? *Analysis*, 67(3):173–180. [6]