What chance-credence norms should not be

Richard Pettigrew

January 14, 2012

A chance-credence norm states how an agent's credences in propositions concerning objective chances ought to relate to her credences in other propositions. The most famous such norm is the Principal Principle (henceforth (PP)), due to David Lewis, which says (roughly) that our credence in a given proposition conditional on the chances being given by a particular probability function ch should be whatever value ch assigns to that proposition ([Lewis, 1980]). However, Lewis noticed that (PP) is inconsistent with many accounts of chance that attempt to reduce chance facts to non-modal facts. More precisely, Lewis showed that (PP) is inconsistent with any account of chance that gives rise to so-called *self-undermining chance functions*—that is, chance functions ch that assign a value of less than 1 to the proposition C_{ch} that the chances are given by ch. And Lewis' favoured account of chance—namely, the Best System Analysis—seems to give rise to just such chance functions.

Against Lewis' pessimistic conclusion, it has been proposed that (PP) is simply the wrong chance-credence norm. It is argued that Lewis was right that we should defer to chance when we set our credences; but he was wrong in the way he formulated what deference requires. Those who defend accounts of chance that give rise to self-undermining chance functions have offered two alternative norms, both of which are consistent with such functions: the first is the New Principle (henceforth (NP)), formulated by Ned Hall and Michael Thau and grudgingly accepted by Lewis ([Hall, 1994], [Thau, 1994], [Lewis, 1994]); and the second is Jenann Ismael's General Recipe, which I will call Ismael's Principle (or (IP)) henceforth ([Ismael, 2008]). However, while (NP) and (IP) are each consistent with self-undermining chances and thus with the sort of reductionism that conflicts with (PP), they are incompatible with one another. Thus, the question arises: Which should the reductionist favour? In this paper, I argue that the consequences of (IP) in the presence of self-undermining chances, while not so bad as the consequences of (PP) in their presence, are nonetheless intolerable. Thus, we ought to favour (NP). I conclude by considering three possible responses to my arguments.

1 The Principal Principle

As Lewis formulated the Principal Principle, it concerned only an agent's credence function at the beginning of her epistemic life—that is, the probability function that gives her credences prior to accumulating any evidence.[1] To state it, we need a little notation. Given a possible world w, the *ur-chance function* at w is the probability function upon which one conditionalizes with the history of w up to a time in order to obtain the chances at w at that time. Given a probability

[1]Throughout, I will assume that our epistemic agents satisfy Probabilism, the norm that demands that credences obey the axioms of the probability calculus.

function ch, let C_{ch} be the proposition *The ur-chances are given by ch*—that is, the proposition that is true at all worlds at which the ur-chance function is ch. Propositions of the form C_{ch} are akin to Lewis' history-to-chance counterfactuals. Thus, we have:

> (PP_0) If b_0 is an agent's initial credence function, then it ought to be the case that, for all propositions A in the algebra $\mathcal{F}$ of propositions about which the agent has an opinion:
> $$b_0(A|C_{ch}) = ch(A)$$
> providing $b_0(C_{ch})$.[2]

We obtain a more general version of the principle, which applies to an agent's credence function at any time in her epistemic life, by appealing to the Bayesian's favoured updating rule, Bayesian Conditionalization, which is stated as follows:

> (BC) If $t < t'$ and b_t and $b_{t'}$ are an agent's credence functions at t and t' respectively and if $K_{t'}$ is the proposition that gives the agent's total evidence at t', then it ought to be the case that:
> $$b_{t'}(A) = b_t(A|K_{t'})$$
> providing $b_t(K_{t'}) > 0$.

Then consider the following general version of (PP_0) (see, e.g., [Nelson, 2009], [Meacham, 2010]):

> (PP) If b_t is an agent's credence function at t and K_t is the proposition that gives the agent's total evidence at t, then it ought to be the case that:
> $$b_t(A|C_{ch}) = ch(A|K_t)$$
> providing $b_t(C_{ch}), ch(K_t) > 0$.

Then we have the following fact:

Proposition 1.1 $(\text{PP}_0) + (\text{BC}) \Rightarrow (\text{PP})$

This proposition provides strong evidence that (PP) is the correct generalisation of Lewis' limited version of the Principal Principle.

However, as Lewis discovered, (PP) (and indeed (PP_0)) is inconsistent in the presence of a certain sort of chance function; and it seems that Lewis' favoured account of chance—namely, the Best System Analysis—gives rise to such chances. According to the Best Systems Analysis, the ur-chances at a given possible world are determined in a certain way by the whole history of non-modal facts that obtain at that world ([Lewis, 1994]). The idea is that the chances at a world are whatever the best theory of that world says they are. A theory of a world consists of a set of propositions about that world and an probability function over some algebra of propositions about that world. The propositions are laws the theory postulates, while the probability function is the ur-chance function of the theory. Other things being equal, the following are virtues of a theory:

[2] Lewis imposed the further constraint (known as Regularity) that $b_0(C_{ch}) > 0$ for all possible chance functions ch. This is not required by anything I wish to say.

- *Simplicity* The simplicity of a theory is measured by the logical complexity of the propositions it contains and their number, as well as by the mathematical and perhaps computational complexity of the chance functions it contains.
- *Strength* The strength of a theory is measured by the proportion of truths about the world that follow from the propositions it contains.
- *Fit* A theory's fit to a world is measured by the ur-chance it assigns to the history of that world. Fit is to an ur-chance function what strength is to the set of laws the theory postulates; it is a measure of how much the theory 'gets right' about the world.

Clearly these properties of simplicity, strength, and fit trade-off against one another. So there may be no unique best theory of a world and perhaps no unique ur-chance function for it. But Lewis is hopeful that this is not the case. Let's grant him his optimism.

Lewis thought that this analysis would give rise to what we call *self-undermining ur-chance functions*. This is a probability function ch such that $ch(C_{ch}) < 1$. They arise whenever ch assigns non-zero probability to the world being very different to the way it actually is; so different, in fact, that if the world were this other way then it would determine the ur-chances to be different from those given by ch.

Why did Lewis think that the Best System Analysis gives rise to self-undermining ur-chances? It is because he thought that, in the case of very simple worlds, in which there is a single chance process that is repeated a fixed number of times, the Best System Analysis assigns the same chance function to each world as is assigned by the frequentist account of chance. Let's take an example: Consider a very simple world that contains only a single coin that will be flipped exactly three times. Suppose it in fact lands TTH. Then, in this very simple case, the Best Systems Analysis will agree with the frequentist, who says that the chance of Heads on each toss is the proportion of Heads amongst all tosses—i.e., $ch(\text{Heads}) = \frac{1}{3}$. But, assuming that the coin tosses are independent, it follows that the chance of the sequence HHH is $(\frac{1}{3})^3 = \frac{1}{27}$. Now, according to the frequentist, in a world in which the sequence of tosses is HHH, the chance of Heads is 1. Let ch' be the chance function at the world HHH—that is, $ch'(\text{Heads}) = 1$. Then, since the world HHH is the only world at which ch' gives the ur-chances, we have that $ch(C_{ch'}) = ch(\text{HHH}) = \frac{1}{27} > 0$. Thus, $ch(C_{ch}) < 1$ since C_{ch} and $C_{ch'}$ are mutually exclusive.

Let us now return to (PP_0) to see why these self-undermining ur-chance functions are so problematic. Suppose ch is one of the probability functions that our agent entertains as a possible ur-chance function—perhaps it is $ch(\text{Heads}) = \frac{1}{3}$ from above. Now consider $b_0(C_{ch}|C_{ch})$. Since b_0 is a probability function, $b_0(C_{ch}|C_{ch}) = 1$. But if b_0 satisfies (PP_0), $b_0(C_{ch}|C_{ch}) = ch(C_{ch}) < 1$. Thus, (PP_0) is incompatible with self-undermining chances.

Thus, Lewis established conclusively that (PP_0) is inconsistent in the presence of self-undermining ur-chance functions. And he believed that the Best System Analysis will, in many worlds including our own, give rise to such functions. Let's assume that Lewis is right about this and consider how the proponent of the Best System Analysis and other reductionist accounts of chance might respond.

2 Alternative chance-credence norms

In response to the inconsistency of (PP) in the presence of self-undermining ur-chance functions, two alternatives to (PP) have been proposed. The first is called the New Principle and it is due to Ned Hall and Michael Thau:

> (NP_0) It ought to be the case that:
>
> $$b_0(A|C_{ch}) = ch(A|C_{ch})$$

Hall's thought is this: Suppose I agree with Lewis that we ought to defer to chance when I fix my initial credences. Then, when I fix my initial credence in A conditional on C_{ch}, I need only consult the ur-chance function ch. But, Hall says, I should not let my *conditional* credence $b_0(A|C_{ch})$ agree with the *unconditional* chance $ch(A)$, as Lewis' Principal Principle demands. Rather, I should let it agree with the *conditional chance* $ch(A|C_{ch})$, since it is my *conditional* credences I am setting ([Hall, 1994]). And (NP_0) generalizes in a very similar way to (PP_0) when we combine it with (BC). Thus, consider:

> (NP) If K_t is the proposition that gives the agent's total evidence at t, then it ought to be the case that:
>
> $$b_t(A|C_{ch}) = ch(A|K_t \wedge C_{ch})$$

Then we have:

Proposition 2.1 (NP_0) + (BC) $\Rightarrow$ (NP)

And we have that (NP) is consistent with self-undermining chances:

Proposition 2.2 *Suppose K_t is a proposition. And suppose $ch_1, \ldots, ch_n$ are probability functions. Suppose further that, for some k, we have:*

- *For every $i \leq k$, C_{ch_i} is consistent with K_t;*
- *For every $i > k$, C_{ch_i} is inconsistent with K_t.*

Now let $\alpha_1, \ldots, \alpha_k$ be non-negative real numbers such that $\sum_{i=1}^{k} \alpha_i = 1$. Then define b_t as follows:

$$b_t(A) = \sum_{i=1}^{k} \alpha_i ch_i(A|K_t \wedge C_{ch_i})$$

Then b_t satisfies (NP) *relative to $ch_1, \ldots, ch_n$.*

The upshot is that (NP) is not only consistent with self-undermining ur-chance functions, but in fact does not narrow down the set of rational credence functions to a small set.

The second alternative to (PP) is Jenann Ismael's (IP). Ismael's thought is this: Like Hall, she agrees with Lewis that an agent ought to defer to chance when she fixes her credences. But, unlike Hall and Lewis, she thinks that deference does not involve fixing one's *conditional* initial credence $b_0(A|C_{ch})$ by appealing to the conditional or unconditional chance of A according to the ur-chance function ch. Rather it involves fixing one's *unconditional* initial credence $b_0(A)$ by appealing to the *unconditional* chances of A according to the various possible ur-chance functions.

But how? When our agent wished to set her credence in A conditional on C_{ch}, she only needed to consult the single ur-chance function ch. Now she needs to consult every possible ur-chance function and then find some way of aggregating the values they assign to A in order to give her credence in A. But of course there is a straightforward way to do this: consider the unconditional chances $ch(A)$ for each possible ur-chance function ch; weight each by the agent's credence $b_0(C_{ch})$ that the ur-chances are given by that function; and sum up the results. Ismael's Principle demands that this is how we set our credence in A. Thus, it demands that an agent's credence in A is her expectation of the ur-chance of A. This gives us the following formulation of the principle as it applies to initial credence functions:

(IP_0) It ought to be the case that:

$$b_0(A) = \sum_{ch} b_0(C_{ch})ch(A)$$

And the natural generalisation is:

(IP) It ought to be the case that:

$$b_t(A) = \sum_{ch} b_t(C_{ch})ch(A|K_t)$$

Again, we have that (IP) is consistent with self-undermining chances:

Proposition 2.3 *Suppose K_t is a proposition. And suppose $ch_1, \ldots, ch_n$ are probability functions. Then there is a probability function b_t such that b_t satisfies* (IP) *relative to $ch_1, \ldots, ch_n$.*

This differs from the analogous result concerning (NP). There, we could give a recipe for constructing credence functions that satisfy the chance-credence norm. Here, since the proof relies on Brouwer's Fixed Point Theorem, which is non-constructive, we must content ourselves with the existence proof. As we will see below, it is sometimes possible to describe functions that satisfy (IP), and indeed sometimes there is a unique such function. But the present proposition tells us only that, in the general case, there is a credence function that satisfies it.

Thus, we have two alternatives to (PP). Both are consistent with self-undermining chances, and indeed with any set of possible chance functions, providing they are probability functions. But (NP) and (IP) are inconsistent with one another in the presence of self-undermining ur-chance functions (though they are equivalent if there are no such functions):

- (NP) $\Leftrightarrow b_t(A) = \sum_{ch} b_t(C_{ch})ch(A|K_t \wedge C_{ch})$
- (IP) $\Leftrightarrow b_t(A) = \sum_{ch} b_t(C_{ch})ch(A|K_t)$

Thus, we must try to choose between them. In the remainder of the paper, we will do just that.

3 Chance-credence norms and updating

Our first contest between (IP) and (NP) concerns their interaction with norms of updating. Bayesians spend a great deal of time discovering, formulating, and justifying norms that govern an agent's credence functions. They spend a great deal less time asking how these various norms interact with one another. Are they, for instance, even mutually consistent? If they make incompatible demands on an agent in certain cases, which should take precedence? In this section, we ask how the chance-credence norms interact with the norms for updating credal states upon receipt of evidence.

We have met the most popular Bayesian updating norm already—it is (BC). We begin here by noting that (IP) and (BC) are incompatible whilst (NP) and (BC) are compatible. That is:

Proposition 3.1 *If b_t satisfies* (NP) *and $b_{t'}$ is obtained from b_t in accordance with* (BC)*, then $b_{t'}$ satisfies* (NP)*.*

However, the analogous result does not hold for (IP).

Proposition 3.2 *There are credence functions b_t and $b_{t'}$ such that b_t satisfies* (IP)*, $b_{t'}$ is obtained from b_t in accordance with* (BC)*, and yet $b_{t'}$ does not satisfy* (IP)*.*

This seems to present a serious objection to (IP). To understand how serious, we must ask: What's so good about Bayesian conditionalization (BC)? One answer that avoids the many problems with Dutch Book arguments is this: Conditionalizing maximises expected epistemic utility for a wide range of putative measures of epistemic utility.[3] More precisely, suppose U is a measure of epistemic utility—that is, it takes a world w and a credence function b and returns a real number $U(b, w)$ that measures the epistemic utility of having b at w. Now say that a credence function b is *immodest* over a set of worlds S if b expects itself to have greater epistemic utility than it expects any other credence function to have when the expectations are calculated over the worlds in the set S: i.e. for all $c \neq b$,

$$\sum_{w \in S} b(w)U(c, w) < \sum_{w \in S} b(w)U(b, w)$$

Immodesty, it is claimed, is a virtue in a credence function, while modesty is a vice. One ought not to have a credence function that expects some other credence function to be better than it expects itself to be.

We have the following result:

Theorem 3.3 (Greaves and Wallace) *If $b_t(\cdot|K_{t'})$ is immodest over some set of worlds that includes $K_{t'}$, then $b_t(\cdot)$ expects $b_t(\cdot|K_{t'})$ to have the greatest epistemic utility, when that expectation is calculated over worlds in $K_{t'}$.*

That is, there is $K \supseteq K_{t'}$ such that, for all $c \neq b_t(\cdot|K_{t'})$

$$\sum_{w \in K_{t'}} b_t(w)U(c, w) < \sum_{w \in K_{t'}} b_t(w)U(b_t(\cdot|K_{t'}), w)$$

This seems to justify (BC): upon receiving evidence $K_{t'}$, we ought to restrict attention to the epistemically possible worlds—i.e. $w \in K_{t'}$—and then ask which credence function will maximize our expected epistemic utility relative to our current credence function and calculated over these

[3] The term 'epistemic utility' is often used to refer to a measure of the epistemic goodness of accepting a scientific hypothesis (see [Maher, 1993]). Here, I use it to mean a measure of the epistemic goodness of adopting a particular credence function. The two notions are clearly related, though it is not clear that all virtues of adopting a scientific theory have natural analogues as virtues of adopting a credence function. For instance, while a scientific hypothesis might unify our evidence, it is not clear that a credence function can.

worlds. Theorem 3.3 tells us that the credence function we should have is the one obtained from b_t by conditionalizing on the evidence $K_{t'}$.

The problem with this justification is that it presupposes that we already know that our epistemic utility function should render $b_t(\cdot|K_{t'})$ immodest. And in order to know that, we must know that $b_t(\cdot|K_{t'})$ is rationally permissible. But it is precisely this that we are trying to discover. After all, (IP) typically renders $b_t(\cdot|K_{t'})$ rationally *im*permissible, since that credence function violates that norm. Thus, we must try a different approach. Rather than assuming that our epistemic utility function renders $b_t(\cdot|K_{t'})$ immodest, let us instead consider a particular epistemic utility function that is plausible independently of its verdict on that credence function. The natural choice is the so-called *Brier score*:

$$B(b, w) := 1 - \sum_{A \in \mathcal{F}} (b(A) - v_w(A))^2$$

where $\mathcal{F}$ is the set of propositions on which b is defined and v_w is the truth-value function at w (that is, $v_w(A) = 1$ if A is true at w; and $v_w(A) = 0$ if A is false at w). The idea is that a credence in a true falsehood is better the closer it is to the maximum credence 1, while a credence in a false proposition is better the closer it is to the minimum credence 0. That is, credence aims at truth; so the epistemic utility of a credence should increase as its distance from the truth-value decreases. Some basic calculus shows that B does indeed render $b_t(\cdot|K_{t'})$ immodest. Thus, if we measure our epistemic utility using B, conditionalizing is the correct way to update upon receipt of new propositional evidence and we vindicate (NP) over (IP).

Nonetheless, the proponent of (IP) should not despair quite yet. The problem with the Brier score B is that it doesn't just make $b_t(\cdot|K_{t'})$ immodest—it makes *all* probability functions immodest. Thus, if B is the correct epistemic utility function, we stand no chance of justifying chance-credence norms by appealing to epistemic utility. This seems unacceptable. It ought to be possible to justify any epistemic norm by appealing to epistemic utility. Indeed, it seems to be part of what we mean by the notion of epistemic utility that any credence functions that are irrational have that irrationality in virtue of some epistemic badness they must have, or can be expected to have. And this badness is exactly what the epistemic utility function is intended to measure.

The natural analogues of the Brier score that nonetheless establish chance-credence norms are the following:

$$C_I^K(b, w) := 1 - \sum_{A \in \mathcal{F}} (b(A) - ch_w(A|K))^2$$

and

$$C_N^K(b, w) := 1 - \sum_{A \in \mathcal{F}} (b(A) - ch_w(A|K \wedge C_{ch_w}))^2$$

If C_I^K is our epistemic utility function, then we value a credence function more the closer it is to the ur-chance function conditional on evidence K. If C_N^K is our epistemic utility function, on the other hand, then we value a credence function more the closer it is to the ur-chance function conditional on evidence K *and the fact that the chances are as they are*.

For these we have:

Proposition 3.4

(i) *Relative to* $C_I^{K_t}$, b_t *is immodest over* K_t *iff* b_t *satisfies* (IP).

(ii) *Relative to* $\mathfrak{C}_N^{K_t}$, b_t *is immodest over* K_t *iff* b_t *satisfies* (NP).

But we also have:

Proposition 3.5

(i) *Relative to* $\mathfrak{C}_I^{K_{t'}}$, b_t *expects the following function to have the greatest epistemic utility, when that expectation is calculated over worlds in* $K_{t'}$*:*

$$b_{t'}(\cdot) := \sum_{ch} b_t(C_{ch}|K_{t'})ch(\cdot|K_{t'})$$

(ii) *Relative to* $\mathfrak{C}_N^{K_{t'}}$, b_t *expects* $b_t(\cdot|K_{t'})$ *to have the greatest epistemic utility, when that expectation is calculated over worlds in* $K_{t'}$.

Thus, consider the following updating norm, which we might call *Ismael Conditionalization*:

> (IC) If $t < t'$ and b_t and $b_{t'}$ are an agent's credence functions at t and t' respectively and if $K_{t'}$ is the proposition that gives the agent's total evidence at t', then it ought to be the case that:
>
> $$b_{t'}(A) = \sum_{ch} b_t(C_{ch}|K_{t'})ch(\cdot|K_{t'})$$

Then the lesson of Propositions 3.4 and 3.5 is this:

- The epistemic utility functions $\mathfrak{C}_I^K$ endorse the chance-credence norm (IP) and the updating norm (IC).
- The epistemic utility functions $\mathfrak{C}_N^K$ endorse the chance-credence norm (NP) and the updating norm (BC).

What's the upshot of all this? It seems to show that it is no objection to Ismael's Principle (IP) that it is incompatible with the usual Bayesian updating norm of conditionalization (BC). After all, the most natural epistemic utility function that endorses (IP) fails to endorse (BC). Rather, it endorses Ismael Conditionalization (IC).

However, this victory is short lived for Ismael's Principle. The problem is that (IP) is not only incompatible with (BC); it is also incompatible with (IC).

Proposition 3.6 *There are credence functions* b_t *and* $b_{t'}$ *such that* b_t *satisfies* (IP), $b_{t'}$ *is obtained from* b_t *in accordance with* (IC), *and yet* $b_{t'}$ *does not satisfy* (IP).

That is, the epistemic utility function that endorses (IP) also endorses an updating norm with which (IP) is incompatible. Herein lies the real problem for (IP).

I conclude this section by considering one apparent way out of this problem. In the face of the incompatibility with (IC), we might weaken (IP) in the following way:

> (IP^-) It ought to be the case that, for each ch, there is α_{ch} such that $\sum_{ch} \alpha_{ch} = 1$ and
>
> $$b_t(A) = \sum_{ch} \alpha_{ch} ch(A|K_t)$$

Then we have that (IP^-) is compatible with (IC):

Proposition 3.7 *If b_t satisfies* (IP^-) *and $b_{t'}$ is obtained from b_t by applying* (IC), *then $b_{t'}$ satisfies* (IP^-).

However, there are problems with this proposal. The first follows from our discussion above: There are credence functions that satisfy (IP^-) that are not immodest relative to the epistemic utility functions that endorse (IC). That is, (IP^-) permits credence functions that expect other credence functions to be better epistemically speaking than they expect themselves to be. The second problem is simply that (IP^-) lacks the intuitive force of (IP). Surely it is irrational to set one's credence in a proposition equal to the weighted sum of the possible chances for that proposition unless the weightings are given by one's credences in the chances.

4 Chance-credence norms and frequencies

Our second contest between (IP) and (NP) concerns their behaviour in the presence of frequentist chances. Lewis seems to accept that the Best System Analysis of chance will agree with the frequentist account in certain sorts of world. Thus, for instance, if a world contains only a single sort of chance event—the tossing of a coin, say—that is repeated a fixed number of times—let's say three times—then, according to the Best System Analysis, the chance of each event of that sort having a particular outcome will be given by the relative frequency of that outcome amongst all instances of the event; and the events will be taken to be independent. Indeed, as we saw above, it is examples like this that lead Lewis to accept that the Best System Analysis will give rise to self-undermining chances. Let us suppose that Lewis is right: in certain simple worlds, the Best System Analysis agrees with the frequentist account. And now let us consider how our rival chance-credence norms function when credences are assigned only to propositions about such worlds.

Throughout, our example will be the one just described: the epistemically possible worlds all consist of exactly three tosses of a coin and nothing more. Thus, they are: HHH, HHT, HTH, HTT, THH, THT, TTH, TTT. And there are four possible ur-chance functions:

$$ch_0(\text{Heads}) = 0 \quad ch_1(\text{Heads}) = \frac{1}{3} \quad ch_2(\text{Heads}) = \frac{2}{3} \quad ch_3(\text{Heads}) = 1.$$

And we have:

$$\begin{aligned}
C_{ch_0} &\equiv \text{TTT} \\
C_{ch_1} &\equiv \text{TTH} \vee \text{THT} \vee \text{HTT} \\
C_{ch_2} &\equiv \text{THH} \vee \text{HTH} \vee \text{HHT} \\
C_{ch_3} &\equiv \text{HHH}
\end{aligned}$$

4.1 Frequencies and non-self-undermining functions

I wish to raise two problems that arise in cases such as this: One concerns what happens when there are some non-self-undermining chance functions that are still epistemically possible for an agent. In our example, the two non-self-undermining chance functions are ch_0 and ch_3. Thus, for an agent whose epistemically possible worlds are those in our example, she will be faced with this situation before the coin has been tossed and she will continue to be in it until two different outcomes have appeared and C_{ch_0} and C_{ch_3} have both been ruled out. This is a point that is often ignored by reductionists about chance: their theories will not always give rise to self-undermining chance functions. The problem is based on the following proposition:

Proposition 4.1 *Suppose ch is non-self-undermining and compatible with evidence K_t. And suppose that $ch'(C_{ch}|K_t) > 0$, for some $ch' \neq ch$. Then, if b_t satisfies* (IP), *then $b_t(C_{ch'}) = 0$.*

That is, Ismael's Principle demands that an agent assign no credence at all those self-undermining chance functions that assign non-zero chance to the chances being given by a non-self-undermining chance function. In our example, this means that the agent must divide her credence over only the non-self-undermining chance functions. She must give no initial credence to C_{ch_1} or to C_{ch_2}. In other words, she must be certain that the coin tosses will be deterministic—they will either always land heads with chance 1, or always land tails with chance 1.

The situation becomes even worse once the coin has been tossed for the first time. It will land H or T. If it lands H, then the only non-self-undermining chance function that remains epistemically possible is ch_3; if it lands T, then only ch_0 remains. Suppose it lands H and our agent learns this. Then C_{ch_0} is no longer epistemically possible for her, but C_{ch_3} is. Moreover, every other remaining chance function—i.e. ch_1 and ch_2—assigns a non-zero probability to C_{ch_3}. So Proposition 4.1 tells us that (IP) demands that our agent be certain of C_{ch_3}. And, similarly, if our agent learns that the first coin toss landed T, (IP) demands that she become certain of C_{ch_0}. Thus, after seeing a single coin toss, our agent must become certain that all remaining coin tosses will go the same way. This, it seems to me, is an unacceptable consequence of (IP).

Of course, while it is easiest to be sure that this result will apply when the chances are given by the relative frequencies, there are also plenty of other situations in which the hypotheses of Proposition 4.1 are satisfied. Non-self-undermining chance functions can arise even for worlds in which the coin toss sometimes lands H and sometimes T. As long as one of the outcomes is rare enough, and providing the overall simplicity of the theory is improved by letting the chance of that outcome be zero, the chance function will not be self-undermining.

How does (NP) fare in this situation? Well, as we saw above, for any ur-chance functions and any evidence, we can divide our credence in any way we please over the ur-chance functions that are compatible with the evidence—whether self-undermining or not—and this gives rise to a credence function over the whole algebra that satisfies (NP). Thus, again, (NP) triumphs.

Before moving on to the second problem with (IP), it is worth noting that, while that norm makes unreasonable demands on an agent in many cases in which non-self-undermining credence functions are still epistemically possible for her, it has a rather attractive consequence when they no longer are. In many cases, when all chance functions that remain are self-undermining, there is a unique credence function that satisfies (IP). Thus, we get a version of Objective Bayesianism based this time on chance-credence norms rather than on something like the Maximize Entropy norm, which is motivated by informational considerations. The following proposition describes two sorts of case in which we always have a unique credence function that satisfies the norm:

Proposition 4.2 *Suppose the remaining epistemically possible ur-chance functions are $ch_1, \ldots, ch_n$. And suppose each ch_i is self-undermining in the presence of K_t—that is, $ch_i(C_{ch_i}|K_t) < 1$. Then, if $n = 2$ or 3, there is a unique credence function b_t that satisfies* (IP) *with respect to these functions.*

Thus, in our coin toss example, if the agent learns that the first two coin tosses are different, then there is a unique credence that the agent should have concerning the outcome of the final coin toss. Perhaps unsurprisingly, it is $b_{t'}(\text{Heads}) = \frac{1}{2}$.

Now suppose we alter our example so that there are exactly four coin tosses, and thus five possible ur-chance functions $ch_0, \ldots, ch_4$. In this case, we have ch_0 and ch_4as the only non-self-undermining functions (where $ch_0(\text{Heads}) = 0$ and $ch_4(\text{Heads}) = 1$). Thus, if our agent learns that the first two coin tosses are different, thereby ruling out C_{ch_0} and C_{ch_4} but leaving C_{ch_1}, C_{ch_2}, and C_{ch_3} as epistemically possible, then there is a unique credence function that satisfies (IP) relative to her body of evidence, since each of ch_1, ch_2, and ch_3 is self-undermining in the presence of that evidence. I leave the question of whether a similar result holds for $n > 3$ as an open problem.

4.2 Frequencies and expected frequencies

We move now to the second problem for (IP) that arises from its behaviour when the chances are given by the relative frequencies of events. The problem is that, in these cases, the chances themselves do not satisfy (IP). This is a feature of frequencies that is not often noted: they don't expect themselves to give the frequencies! Consider our case from above. Then, though we have:

$$ch_0(A) = \sum_{i=0}^{3} ch_0(C_{ch_i})ch_i(A)$$
$$ch_3(A) = \sum_{i=0}^{3} ch_3(C_{ch_i})ch_i(A)$$

for all A. We also have

$$ch_1(\text{HHH}) = \frac{1}{27} \neq \frac{29}{243} = \sum_{i=0}^{3} ch_1(C_{ch_i})ch_i(\text{HHH})$$

and

$$ch_2(\text{HTH}) = \frac{4}{27} \neq \frac{20}{243} = \sum_{i=0}^{3} ch_2(C_{ch_i})ch_i(\text{HTH})$$

Thus, while ch_0 and ch_3 satisfy (IP)—and, in general, all non-self-undermining ur-chances functions will—ch_1 and ch_2 do not.

Why is this a problem? First, it might seem that, if we defer to the chances, it ought to be rational to have a credence function that exactly matches them. But this is ruled out as irrational by (IP) when the chances are given by frequencies. Second, the following meta-normative principle seems plausible: It is not rational to defer to an epistemic expert that does not defer to itself. Our chance-credence norms are intended to give a precise formulation of the claim that we ought to defer to the chances. Thus, if the chances themselves don't satisfy that norm, then they don't defer to themselves, and our principle tells us that we've formulated it incorrectly.

As before, (NP) does not face the same problem, *at least when the chances are given by frequencies in this way*. The crucial feature of frequencies is this: for each chance function ch_i and each chance hypothesis C_{ch_j}, if C_{ch_j} is the disjunction of the following possible worlds $w_1, \ldots, w_k$, then $ch_i(w_m) = ch_i(w_n)$. This is so because, whenever two worlds have the same chance function, they have the same distribution of outcomes of the chance events and thus any chance function will assign each of these distributions the same chance. For example, consider our example from above of worlds that contain only three coin tosses. Then

$$C_{ch_1} \equiv \mathrm{HTT} \vee \mathrm{THT} \vee \mathrm{TTH}$$

And each ch_i will assign the same chance to each world HTT, THT, and TTH. For instance, ch_0 will assign each world 0; ch_1 will assign each world $(\frac{1}{3})(\frac{2}{3})^2 = \frac{4}{27}$; and so on. Together with the following proposition, this guarantees that, whenever the chances are given by frequencies, they will satisfy (NP):

Proposition 4.3 *Suppose the possible ur-chance functions are $ch_0, \ldots, ch_n$. Suppose that for all ch_i and all worlds $w, w' \in C_{ch_i}$, we have $ch_j(w) = ch_j(w')$, for all ch_j. Then each possible ur-chance function satisfies* (NP).

Thus, once again, we have a reason to favour (NP) over (IP), though this time it is qualified, for it may be that chance functions for more complicated worlds, which are not simply given by frequences, do not satisfy (NP). For it is certainly not true that just any chance function does so.

5 Replies

In the previous two sections, we considered some of the unpalatable consequences of (IP) and showed that (NP) does not share them. I conclude by considering some possible responses to the arguments given above.

The first response is based on a claim that Ismael makes in the paper in which she introduces (IP). She maintains that chance functions are usually not defined at propositions giving chance hypotheses. Thus, if *ch* is a possible ur-chance function, it will typically not be defined at propositions of the form $C_{ch'}$. If this is the case, then the arguments above fail, since all turn in one way or another on assuming that the following is a consequence of (IP):

$$b_t(C_{ch'}) = \sum_{ch} b_t(C_{ch}) ch(C_{ch'} | K_t)$$

for each ch'. But, unfortunately, for those who wish to defend Lewis' Best System Analysis and indeed for reductionists of many other stripes, this response cannot work. To see this, recall the brief description of the Best System Analysis from above. Other things being equal, a theory is better, according to this account, the greater the chance it assigns to the *history of the actual world*. This is the so-called 'fit' of the theory. Thus, the chance functions that the theory posits must be defined on this history. And if this is the case, then there seems every reason to believe that they will be defined on other possible, non-actual histories. But of course, according to the Best System Analysis, a chance hypothesis is simply a disjunction of possible histories—it is the disjunction of those possible histories whose best theory sets the chances equal to those posited by the hypothesis. Thus, in our example of the previous section, the chance hypothesis C_{ch_1} is equivalent to the disjunction TTH $\vee$ THT $\vee$ HTT. So it seems that, in order for the Best System Analysis to work, chances must be defined for chance hypotheses, as required by our arguments above.

The second response is that, since (IP) is based on the most natural notion of deference—more natural, it might be claimed, than the original (PP)—and since it has been shown to have unacceptable consequences in the presence of self-undermining chance functions, our conclusion should not be that (NP) is the correct norm, but rather that self-undermining chances are, as Lewis feared, unacceptable. While tempting, I think this response is too quick. We have the intuition that we should use the chance functions to fix our credences. We also have the further intuition that we should do this by treating the chances as some sort of epistemic expert to which we defer. What we do not have is an intuition that picks out a particular precise way of formulating what this deference requires. (IP) and (NP) present two alternative formulations. Perhaps we are drawn more strongly to one than to the other. But this intuitive appeal is not enough to lead us to reject a certain account of chance when we discover that that account is incompatible with the notion of deference that we intuitively favoured. It would be enough if the precise formulation of deference given by (NP) were simply unacceptable. But I take it that Hall's motivation that I described above renders it at least acceptable, if not initially preferable. Thus, we would do better to conclude, in the face of the arguments given here, that (NP) is the correct chance-credence norm, rather than that the Best System Analysis fails.

6 Conclusion

In sum: Lewis' favoured account of chance—namely, the Best System Analysis—gives rise to self-undermining ur-chance functions. These are inconsistent with Lewis' favoured chance-credence norm—namely, the Principal Principle. In order to save the Best System Analysis and the intuition that we ought to defer to chance in fixing our credences, two alternative chance-credence norms have been formulated in an attempt to say what is really required by such deference. These are (NP) and (IP). I have argued that the former is to be preferred over the latter for two reasons: First, (IP) is incompatible with the only two reasonable updating norms in the vicinity—namely, (BC) and (IC)—while (NP) is compatible with (BC); and second, (IP) has unpalatable consequences when chances are given by frequencies, as they would be in very simple worlds containing a single chance process, while (NP) avoids these consequences.

7 Proofs

Proof of Proposition 1.1. Suppose (PP_0) and (BC). Then $b_t(\cdot) = b_0(\cdot|K_t)$. Thus,

$$b_t(A|C_{ch}) = b_0(A|C_{ch} \wedge K_t) = \frac{b_0(A \wedge K_t|C_{ch})}{b_0(K_t|C_{ch})} = \frac{ch(A \wedge K_t)}{ch(K_t)} = ch(A|K_t)$$

as required. □

Proof of Proposition 2.1. Exactly analogous to the proof of Proposition 1.1. □

Proof of Proposition 2.2. Define b_t as in the theorem. Then

$$\begin{aligned} b_t(A|C_{ch_j}) &= \frac{b_t(A \wedge C_{ch_j})}{b_t(C_{ch_j})} \\ &= \frac{\sum_{i=1}^{k} \alpha_i ch_i(A \wedge C_{ch_j}|K_t \wedge C_{ch_i})}{\sum_{i=1}^{k} \alpha_i ch_i(C_{ch_j}|K_t \wedge C_{ch_i})} \\ &= \frac{\alpha_j ch_j(A|K_t \wedge C_{ch_j})}{\alpha_j} \\ &= ch_j(A|K_t \wedge C_{ch_j}) \end{aligned}$$

as required. □

Proof of Proposition 2.3. Any fixed point of the following function from probability functions to probability functions will satisfy (IP):

$$b_t(\cdot) \mapsto \sum_{ch} b_t(C_{ch}) ch(\cdot|K_t)$$

By Brouwer's Fixed Point Theorem, the function has a fixed point. □

Proof of Proposition 3.1. Analogous to proof of Proposition 1.1 above. □

Proof of Proposition 3.2. We know from Proposition 4.2 (proved below) that, when the epistemically possible ur-chance functions are $ch_1, \ldots, ch_n$ and $n = 2$ or 3 and each ch_i is self-undermining even in the presence of evidence K_t—that is, $ch_i(C_{ch_i}|K_t) < 1$—then there is a unique credence function that satisfies (IP). Now consider the case in which each world contains exactly four coin tosses and nothing more. So the possible chance functions are $ch_0, \ldots, ch_4$. Now suppose that our evidence at t is $K_t \equiv$ HT—that is, we know that the first coin toss was H and the second T. Then we are in a situation in which there are three epistemically possible ur-chance functions—namely, ch_1, ch_2, and ch_3—and each is self-undermining in the presence of K_t. Thus, we know that there is a unique b_t such that b_t satisfies (IP). Moreover, we can calculate this b_t. It is determined by the following values:

$$b_t(C_{ch_1}) = \frac{2}{7} \quad b_t(C_{ch_2}) = \frac{3}{7} \quad b_t(C_{ch_3}) = \frac{2}{7}$$

Now suppose that our evidence at t' is $K_{t'} \equiv$ HTH. Then we are in a situation in which there are two epistemically possible ur-chance functions—namely, ch_2 and ch_3—and each is self-undermining in the presence of $K_{t'}$. Again, we can calculate the unique $b_{t'}$ that satisfies (IP). It has:

$$b_{t'}(C_{ch_2}) = \frac{1}{3} \quad b_{t'}(C_{ch_3}) = \frac{2}{3}$$

But we can also calculate $b_t(\cdot|\text{HTH})$. For instance, we have:

$$b_t(C_{ch_2}|\text{HTH}) = \frac{2}{5}$$

Thus, we have

$$b_t(C_{ch_2}|\text{HTH}) = \frac{2}{5} \neq \frac{1}{3} = b_{t'}(C_{ch_2})$$

Thus, the probability function obtained from b_t by conditionalizing on HTH disagrees with the unique credence function that satisfies (IP) relative to this evidence. Thus, (IP) is not preserved by (BC), as required. □

Proof of Proposition 3.6. We use the same example as in the proof of Proposition 3.2. Let's consider the credence function c obtained from b_t by updating using (IC) on proposition HTH. We have:

$$\begin{aligned} c(C_{ch_2}) &= b_t(C_{ch_2}|\text{HTH})ch_2(C_{ch_2}|\text{HTH}) + b_t(C_{ch_3}|\text{HTH})ch_3(C_{ch_2}|\text{HTH}) \\ &= \left(\frac{2}{5}\times\frac{1}{2}\right) + \left(\frac{3}{5}\times\frac{9}{36}\right) = \frac{7}{20} \end{aligned}$$

Thus,

$$c(C_{ch_2}) = \frac{7}{20} \neq \frac{1}{3} = b_{t'}(C_{ch_2})$$

Thus, the probability function obtained from b_t by updating in accordance with (IC) on HTH disagrees with the unique credence function that satisfies (IP) relative to this evidence. Thus, (IP) is not preserved by (BC), as required. □

Proof of Proposition 3.3. Suppose $b_t(\cdot|K_{t'})$ is immodest over $K \supseteq K_{t'}$. Then we have that, for all $c \neq b_t(\cdot|K_{t'})$,

$$\begin{aligned} \sum_{w\in K} b_t(w|K_{t'})U(c,w) &< \sum_{w\in K} b_t(w|K_{t'})U(b_t(\cdot|K_{t'}),w) \\ \sum_{w\in K_{t'}} \frac{b_t(w)}{b_t(K_{t'})}U(c,w) &< \sum_{w\in K_{t'}} \frac{b_t(w)}{b_t(K_{t'})}U(b_t(\cdot|K_{t'}),w) \\ \sum_{w\in K_{t'}} b_t(w)U(c,w) &< \sum_{w\in K_{t'}} b_t(w)U(b_t(\cdot|K_{t'}),w) \end{aligned}$$

as required. □

Proof of Proposition 3.4.

(i) This follows from the following fact:

$$0 = \frac{d}{dx}\sum_w b(w)(x - ch_w(A|K))^2 = 2\sum_w b(w)(x - ch_w(A|K))$$

if, and only if,

$$x = \frac{\sum_w b(w)ch_w(A|K)}{\sum_w b(w)} = \sum_w b(w)ch_w(A|K) = \sum_{ch} b(C_{ch})ch(A|K)$$

(ii) The proof is exactly analogous.

□

Proof of Proposition 3.5. The proof is exactly analogous to the proof of Proposition 3.4. □

Proof of Proposition 3.7. Straightforward. □

Proof of Proposition 4.1. Suppose $ch(C_{ch}) = 1$. And suppose b_t satisfies (IP). Then

$$b_t(C_{ch}) = \sum_{ch'} b_t(C_{ch'})ch'(C_{ch}|K_t)$$

So

$$0 = \sum_{ch' \neq ch} b_t(C_{ch'})ch'(C_{ch}|K_t)$$

Since $b_t(C_{ch'}), ch'(C_{ch}|K_t) \geq 0$ for all ch, ch', if $ch'(C_{ch}|K_t) > 0$ for $ch' \neq ch$, then $b_t(C_{ch'}) = 0$, as required. □

Proof of Proposition 4.2. A credence function b_t satisfies (IP) iff

$$\begin{pmatrix} ch_1(C_{ch_1}|K_t) & \cdots & ch_n(C_{ch_1}|K_t) \\ \vdots & & \vdots \\ ch_1(C_{ch_n}|K_t) & \cdots & ch_n(C_{ch_n}|K_t) \end{pmatrix} \begin{pmatrix} b_t(C_{ch_1}) \\ \vdots \\ b_t(C_{ch_n}) \end{pmatrix} = \begin{pmatrix} b_t(C_{ch_1}) \\ \vdots \\ b_t(C_{ch_n}) \end{pmatrix}$$

Writing c_{ij} for $ch_j(C_{ch_i}|K_t)$ and x_i for $b(C_{ch_i})$, we have:

$$\begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

And this is equivalent to:

$$\begin{pmatrix} c_{11}-1 & c_{12} & \cdots & c_{1(n-1)} & c_{1n} \\ c_{21} & c_{22}-1 & \cdots & c_{2(n-1)} & c_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ c_{(n-1)1} & c_{(n-1)2} & \cdots & c_{(n-1)(n-1)}-1 & c_{(n-1)n} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Let C be the $n \times n$ matrix on the left-hand side of the equation. So we have that there is a unique b_t that satisfies (IP) in this situation iff C is invertible iff $\det(C) \neq 0$.

If $n = 2$, then $\det(C) = c_{11} - c_{12} - 1 = -(c_{12} + c_{21}) = 0$ iff $c_{12} = 0 = c_{21}$ iff $c_{11} = 1 = c_{22}$ iff $ch_1(C_{ch_1}|K_t) = 1$ and $ch_2(C_{ch_2}|K_t) = 1$ iff ch_1 and ch_2 are non-self-undermining in the presence of K_t. Thus, if ch_1 and ch_2 are self-undermining in the presence of K_t, we have $\det(C) \neq 0$ and there is a unique b_t. It is straightforward to specify this b_t. It is this:

$$b_t(C_{ch_1}) = \frac{ch_2(C_{ch_1}|K_t)}{ch_2(C_{ch_1}|K_t) + ch_1(C_{ch_2}|K_t)} \qquad b_t(C_{ch_2}) = \frac{ch_1(C_{ch_2}|K_t)}{ch_2(C_{ch_1}|K_t) + ch_1(C_{ch_2}|K_t)}$$

If $n = 3$, then $\det(C) = c_{21}(c_{32} + c_{23}) + c_{32}(c_{12} + c_{32} + c_{33}) + c_{12}c_{23} + c_{13}(c_{12} + c_{32} + c_{21})$. Since $ch_i(C_{ch_i}) < 1$, for $i = 1, 2, 3$, we have that $c_{11}, c_{22}, c_{33} < 1$. And, since $c_{1i} + c_{2i} + c_{3i} = 1$, we have $c_{12} + c_{32} = 1 - c_{11} > 0$. Thus, if $\det(C) = 0$, we must have $c_{13}, c_{32} = 0$. But this means that we cannot have $c_{23} = 0$ nor $c_{12} = 0$. And thus $c_{12}c_{23} > 0$, which gives a contradiction. Thus, again, we have that there is a unique credence function that satisfies (IP). □

Proof of Proposition 4.3. Suppose that, for all i, j, if $w, w' \in C_{ch_i}$, we have $ch_j(w) = ch_j(w')$. Then, if $w \in C_{ch_j}$, we have

$$ch_i(w|C_{ch_j}) = \frac{ch_i(w)}{ch_i(C_{ch_j})} = \frac{ch_i(w)}{\sum_{w' \in C_{ch_j}} ch_i(w')} = \frac{1}{|C_{ch_j}|}$$

Thus,

$$ch_i(A|C_{ch_j}) = \frac{|A|}{|C_{ch_j}|} = ch_j(A|C_{ch_j})$$

so ch_i satisfies (NP), as required. □

References

[Hall, 1994] Hall, N. (1994). Correcting the Guide to Objective Chance. *Mind*, 103:505–518.

[Ismael, 2008] Ismael, J. (2008). Raid! Dissolving the Big, Bad Bug. *Noûs*, 42(2):292–307.

[Lewis, 1980] Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*, volume II. University of California Press, Berkeley.

[Lewis, 1994] Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, 103:473–490.

[Maher, 1993] Maher, P. (1993). *Betting on Theories*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press, Cambridge.

[Meacham, 2010] Meacham, C. J. G. (2010). Two Mistakes Regarding the Principal Principle. *British Journal for the Philosophy of Science*, 61(2):407–431.

[Nelson, 2009] Nelson, K. (2009). On Background: Using Two-Argument Chance. *Synthese*, 166:165–186.

[Pettigrew, 2012] Pettigrew, R. (2012). Accuracy, Chance, and the Principal Principle. *Philosophical Review*, 121(2):241–275.

[Thau, 1994] Thau, M. (1994). Undermining and Admissibility. *Mind*, 103:491–504.