

★

The Principal Principle implies the Principle of Indifference

★

James Hawthorne, Jürgen Landes, Christian Wallmann and Jon Williamson

Draft of November 21, 2014

Abstract

We argue that David Lewis' Principal Principle implies a version of the Principle of Indifference. The same is true for similar principles which need to appeal to the concept of admissibility. Such principles are thus in accord with objective Bayesianism, but in tension with subjective Bayesianism.

Keywords: Bayesianism, Principal Principle, Calibration, Principle of Indifference, objective Bayesianism, subjective Bayesianism.

§1

The Argument

Lewis (1980) put forward the following principle as a constraint on a reasonable initial credence function P , which is taken to be a probability function:

Principal Principle. $P(A|XE) = x$, where X says that the chance at time t of proposition A is x and E is any proposition that is compatible with X and admissible at time t .

The Principal Principle implies that, if one's evidence includes the proposition that the chance at time t of A is x then one should believe A to degree x , as long as one's other evidence E does not include anything that defeats this ascription of rational belief. If $x < 1$ and E logically entails A then E is a defeater, for instance, since by the laws of probability, $P(A|XE) = 1 \neq x$. On the other hand, if E is a proposition entirely about matters of fact no later than time t then as a rule E is admissible and not a defeater (Lewis, 1980, pp. 92–96). It is also intuitively plausible that the following two conditions hold, where, henceforth, F is any proposition that is contingent (neither necessarily true nor necessarily false) and atomic (not logically complex).¹

Condition 1. If E is not a defeater and XE contains no information that renders F relevant to A , then EF is not a defeater.

Condition 2. If E is not a defeater and XE contains no information relevant to F , then $E(A \leftrightarrow F)$ is not a defeater.

Consider an example in which there are two tests, an a -test and an f -test, each of which has two possible possible outcomes, positive or negative. You are told just

that the chance of the a -test yielding a positive result is 0.8 (X). The Principal Principle implies that you should believe the a -test turns out positive to degree 0.8, $P(A|X) = 0.8$. By Condition 1, learning that the f -test yields a positive outcome (F) should not defeat this application of the Principal Principle, $P(A|XF) = 0.8$, because there is no evidence linking F to A in this example. By Condition 2, learning that the a -test is positive if and only if the f -test is positive should not defeat the Principal Principle, $P(A|X(A \leftrightarrow F)) = 0.8$, because, although $A \leftrightarrow F$ specifies a link between F and A , there is no evidence concerning F here.

We shall take the claim that E is not a defeater to hold just when $P(A|XE) = x = P(A|X)$. Moreover, we shall take the supposition that XE contains no information that renders F relevant to A to imply that $P(A|FXE) = P(A|XE)$. Then Condition 1 provably holds:

Proposition 1. Suppose E is a non-defeater and XE contains no information that renders F relevant to A . Then EF is a non-defeater.

Proof: $P(A|FXE) = P(A|XE) = P(A|X)$ since E is a non-defeater. □

It turns out that, under certain conditions, one should believe F to degree 1/2 under evidence XE :

Proposition 2. Suppose that E , EF and $E(A \leftrightarrow F)$ are non-defeaters and that $0 < x < 1$. Then the Principal Principle implies that $P(F|XE) = 1/2$.

Proof: Since both EF and $E(A \leftrightarrow F)$ are non-defeaters,

$$P(A|FXE) = P(A|(A \leftrightarrow F)XE).$$

That these conditional probabilities are well defined implies that their conditions have non-zero probability and so $P(F|XE), P(A \leftrightarrow F|XE) > 0$. Applying Bayes' theorem,

$$\frac{P(F|AXE)P(A|XE)}{P(F|XE)} = \frac{P(A \leftrightarrow F|AXE)P(A|XE)}{P(A \leftrightarrow F|XE)}.$$

Now $P(F|AXE) = P(A \leftrightarrow F|AXE)$, so the numerators are equal. Each side of the above equation is equal to $P(A|FXE) = x > 0$ so the numerators are non-zero. Therefore the denominators are equal,

$$P(F|XE) = P(A \leftrightarrow F|XE).$$

Hence,

$$\begin{aligned} P(F|XE) &= P(FA|XE) + P(\neg F \neg A|XE) \\ &= P(A|FXE)P(F|XE) + (1 - P(F|\neg AXE))P(\neg A|XE) \\ &= xP(F|XE) + \left(1 - \frac{P(\neg A|FXE)P(F|XE)}{P(\neg A|XE)}\right)(1 - x) \\ &= xP(F|XE) + (1 - x) - P(F|XE)(1 - x). \end{aligned}$$

(Note in the above that $P(\neg A|XE) = 1 - x > 0$.) Collecting terms, $2(1 - x)P(F|XE) = 1 - x$, so $P(F|XE) = 1/2$. □

Now suppose that E is a non-defeater and XE contains no information relevant to F or that renders F relevant to A . By Conditions 1 and 2, neither EF nor $E(A \leftrightarrow$

F) are defeaters. Hence by Proposition 2, $P(F|XE) = 1/2$. But this is a version of the Principle of Indifference, since it says that, given a suitable lack of information about F , one should believe F and $\neg F$ to exactly the same degree, under evidence XE . Indeed, since XE contains no information relevant to F , $P(F|XE) = P(F)$, so $P(F) = 1/2$ too. Thus the Principle of Indifference also holds for unconditional initial credences.

★

Although there is some debate about how the Principal Principle is best formulated, it is perhaps fair to say that most philosophers who are concerned with norms on rational degree of belief endorse a version of the Principal Principle. On the other hand, the Principle of Indifference is still viewed largely with suspicion. The above considerations suggest that the Bayesian epistemologist should either embrace both principles, in line with objective Bayesianism (see, e.g., Williamson, 2010), or deny both principles, in line with radical subjectivism. Either way, the Principal Principle as a half-way house between radically subjective Bayesianism and objective Bayesianism becomes an unstable position.

Interestingly, this line of argument does not depend on the structure of the proposition X . Exactly the same considerations apply to any principle that seeks to constrain rational credence on the basis of some feature \mathcal{F} of a proposition X , in the presence of a proposition E that is a non-defeater:

Superprincipal Principle. $P(A|XE) = x$, where X relates A and x by feature \mathcal{F} , and E is any proposition that is compatible with X and admissible with respect to \mathcal{F} .

Other examples of the Superprincipal Principle include other formulations of the Principal Principle, the Reflection Principle (where X says that one's rational credence in A at some future time t is x), and testimony principles (where X says that an appropriate authority on A believes A to degree x).

The general point to be made is this. If one wants to depart from radical subjectivism by constraining certain credences via some instance of the Superprincipal Principle, then one needs to invoke the notion of defeat—i.e., one needs to consider whether proposition E is compatible and admissible. Conditions 1 and 2 must hold because these conditions encapsulate core intuitions about defeat. But then one needs to move to objectivism insofar as one needs to accept applications of the Principle of Indifference that take the form: if E is a non-defeater and XE contains no information relevant to F or that renders F relevant to A , then $P(F|XE) = 1/2$.

Note that some versions of the Principal Principle avoid explicit appeal to the notion of defeat. For example, Meacham (2010, p. 426) maintains that, if G grounds chance in the sense that it determines that the chance of A is x , then $P(A|G) = x$. However, one still needs to consider defeat in practice. This is because one needs to be able to decide questions to do with grounding: one needs to decide whether G determines that the chance of A is x , in order to apply this version of the Principal Principle. Suppose G is XE where it is apparent that grounding proposition X determines that the chance of A is x , and E is some other proposition. In order to decide whether G determines that the chance of A is x , one needs to ascertain whether E is a defeater in this context—i.e., whether XE grounds a different chance to that which is grounded by X . Conditions 1 and 2 then become pertinent and the Principle of Indifference follows by the argument presented above, except with

X interpreted as grounding a chance claim rather than directly expressing it. So, while some versions of the Principal Principle are not explicitly instances of the Superprincipal Principle as formulated above, they do implicitly need to appeal to defeat and hence also lead to the Principle of Indifference.

§2

Some Objections Met

One might try to undermine the above argument by endorsing the Principal Principle while rejecting Condition 2.

One possible objection to Condition 2 proceeds by noting that there are similar intuitively plausible principles that cannot hold:

Condition 3. If E is not a defeater and XE contains no information relevant to F , then $E(A \rightarrow F)$ is not a defeater.

Condition 4. If E is not a defeater and XE contains no information relevant to F , then $E(F \rightarrow A)$ is not a defeater.

As we shall see next, Condition 3, respectively Condition 4, can only hold in general if $P(A \rightarrow F|XE) = 1$, respectively $P(F \rightarrow A|XE) = 1$, which is implausible when there is no information relevant to F and the only information relevant to A is the chance information X .

Proposition 3. Suppose that neither E nor $E(A \rightarrow F)$ are defeaters and that $0 < x < 1$. Then the Principal Principle implies that $P(A \rightarrow F|XE) = 1$. Equally, if $E(F \rightarrow A)$ is no defeater, then $P(F \rightarrow A|XE) = 1$.

Proof: Note that $(A \rightarrow F) \equiv \neg(A \neg F)$. We have

$$x = P(A|XE) = P(A|XE \neg(A \neg F))P(\neg(A \neg F)|XE) + P(A|XE A \neg F)P(A \neg F|XE).$$

Now, since $P(A|XE(A \neg F)) = 1$, we have

$$x = xP(\neg(A \neg F)|XE) + 1 - P(\neg(A \neg F)|XE).$$

Hence, $x - 1 = (x - 1)P(\neg(A \neg F)|XE)$ and therefore, $1 = P(\neg(A \neg F)|XE)$. The proof of the second claim is similar. \square

If Conditions 3 and 4 are as plausible as Condition 2, but cannot hold, then this diminishes the case for Condition 2.

However, the intuitive appeal of Conditions 3 and 4 is illusory. The proposition $F \rightarrow A$ rules out all the $F \neg A$ worlds, and so provides information to favour A over $\neg A$. Hence it is by no means clear that $E(F \rightarrow A)$ should not be a defeater. Similarly, $A \rightarrow F$ rules out the $A \neg F$ worlds, and so favours $\neg A$ over A . But Condition 2 is not susceptible to this problem: $A \leftrightarrow F$ favours neither A nor $\neg A$ since it rules out A and $\neg A$ worlds in equal measure.

While this consideration may successfully undermine Conditions 3 and 4, it does point to another possible objection to Condition 2. What if the A and $\neg A$ worlds that are ruled out by $A \leftrightarrow F$ do not have equal measure, in terms of subjective probability? If one gives higher prior probability to $\neg A F$ than one does to $A \neg F$, then $A \leftrightarrow F$ does apparently favour A over $\neg A$. Similarly, if $\neg A F$ has lower prior

probability than $A \neg F$ then $A \leftrightarrow F$ apparently favours $\neg A$ over A . Thus it appears that $E(A \leftrightarrow F)$ should be a defeater.

In defence of Condition 2, three points are pertinent. First, the Principal Principle (or indeed any other instance of the Superprincipal Principle) constrains *initial* credences—there is no prior-to-initial credence function that can play a role. Second, with any such principle it is important that whether a proposition E is a defeater depends on characteristics of the proposition itself and its relation to X and A , not merely on one's initial credence function, for otherwise the principle trivialises to: $P(A|XE) = x$ unless $P(A|XE) \neq x$. Thus Lewis was careful to characterise defeat in terms of compatibility and admissibility, with admissibility depending very much on the nature of the proposition E . Third, the antecedent of Condition 2 ensures that $E(A \leftrightarrow F)$ provides no evidential grounds to prefer $\neg AF$ over $A \neg F$ or vice versa; any such preference is entirely arbitrary. Therefore, Condition 2 remains plausible: there is nothing in $E(A \leftrightarrow F)$ to defeat an application of the Principal Principle. Applying Proposition 2, Condition 2 precludes a subjective weighting that favours one of $\neg AF$ and $A \neg F$ over the other, not the other way round. Evidential considerations trump arbitrary subjective choice when determining whether a proposition is a defeater.²

A third possible objection to Condition 2 is that it fails under a definition of admissibility put forward by Meacham (2010, p. 418). Meacham suggests that E is admissible if and only if XE is logically equivalent to some disjunction $T_1H_1 \vee \dots \vee T_kH_k$, where each T_i is a complete chance theory that implies X , i.e., that the chance of A at time t is x , and each H_i is a complete history up to time t of a world at which T_i holds. In particular, any proposition not entailed by such a chance-history disjunction is inadmissible. Now suppose that the chance of A is non-trivial, i.e., $0 < x < 1$. Then A must be a proposition about the future and so must $A \leftrightarrow F$. As Hume noted, the future is logically independent of the past; thus $A \leftrightarrow F$ is inadmissible. According to Meacham's definition of admissibility, then, Condition 2 systematically fails.

Note, however, that this definition of admissibility is extraordinarily restrictive. If F is any contingent atomic proposition about the future then it is deemed inadmissible. For example, the proposition *the sun will rise tomorrow* (F) is deemed to defeat an application of the Principal Principle with respect to an unrelated proposition *it will rain this evening* (A). This is not only contrary to intuition—it is hard to see how F can trump the chance of A in determining a reasonable initial credence in A —but it also violates Condition 1, which, we argued, provably holds. Thus this objection to Condition 2 throws the baby out with the bath water.

Having defended Condition 2 against three possible objections, we shall see next that, even if some new objection is found that tells against Condition 2, this will not necessarily undermine the main conclusion of §1, namely that the Principal Principle (or any other instance of the Superprincipal Principle) is in tension with subjective Bayesianism.

Suppose that some pathological F 's can be found that do not conform to Condition 2—perhaps propositions that, although contingent and atomic, are self-informative in the sense that such an F imposes some constraint on its own probability.³ Even if one rejects Condition 2 as a general principle, it remains plausible that at least some restricted version of Condition 2 will hold. I.e., it remains plausible that there is some natural subclass of propositions F (non-self-informative propositions, perhaps) such that $E(A \leftrightarrow F)$ are non-defeaters, when E is a non-defeater and XE contains no information relevant to F . However, for each such non-defeater $E(A \leftrightarrow F)$, Proposi-

tion 2 forces $P(F|XE) = 1/2$. This leads to a version of the Principle of Indifference, albeit a qualified version. That it implies any such Principle of Indifference is enough to put the Principal Principle in conflict with subjectivism, i.e., for our main line of argument to go through.

On the other hand, to hold that Condition 2 fails more routinely—not just in pathological cases—would indicate that the Principal Principle admits so many counterintuitive defeaters as to render the principle itself unviable.⁴

Acknowledgements

We are very grateful to Seamus Bradley, David Corfield, Luke Fenton-Glynn, Jan-Willem Romeijn, Gregory Wheeler and two anonymous referees for helpful comments. Jürgen Landes and Jon Williamson were supported by the UK Arts and Humanities Research Council, on the project *From objective Bayesian epistemology to inductive logic*. Christian Wallmann was supported by the Federal Ministry of Science, Research and Economy of the Republic Austria (BMWFW) in cooperation with the Austrian Agency for International Mobility and Cooperation in Education, Science and Research (OeAD-GmbH) (Grant: Marietta Blau).

Notes

¹This restriction to contingent atomic propositions is required to ensure that F does not provide information relevant to its own probability. The argument of this section does not go through in general for propositions that contain such information. Examples of self-informative propositions include the following. Necessarily true and false propositions should arguably be given credence 1 and 0 respectively, even if they are not logically true, respectively logically false. A logically complex proposition such as *ticket number 97 won a fair thousand-ticket lottery with one winner* should arguably be believed to degree 1/1000 or thereabouts, in the absence of other information. A logically complex proposition that is a conjunction of very many atomic propositions should arguably be given low credence, in proportion to the number of conjuncts, in the absence of other information. In such cases, Condition 1 and Condition 2 can fail. In the last case, for instance, Condition 2 cannot apply since otherwise if we take as F s the logically complex propositions BC , $B\neg C$, $\neg BC$ and $\neg B\neg C$, where B and C are atomic propositions, then Proposition 2 would have their probabilities sum up to 2, instead of the value 1 forced by the axioms of probability.

²The subjectivist might want to dig in here by insisting that initial credences, rather than features of the propositions in question, decide defeat. (Joyce (2010, pp. 300–301) appears to adopt this sort of position, though from the perspective of imprecise probability.) The subjectivist might argue that, although such a view trivialises the Principal Principle, it still sheds some light on the role of chance, and therefore serves a useful purpose. In response, note that the Principal Principle is formulated in terms of compatibility and admissibility because it is expected to do both things: to shed light on chance *and* to provide a substantive constraint on credence. If one were to disavow this second goal and develop a version of the Principal Principle that offers no substantive constraint on credence, one would be left with radical subjectivism. Thus our main conclusion would still go through—such a position would fail to offer a viable middle ground between radical subjectivism and objectivism. The same point can be made with regard to other instances of the Superprincipal Principle, which are also intended as substantive constraints on credence.

³Consider, for example, the contingent atomic proposition *Fido is exploding*, where Fido refers to an unseen dog. One might think that this proposition is self-informative, thereby violating Condition 2, on the grounds that once one understands this proposition it is plain that it should be given very little credence. Contrary to appearances, however, this proposition is not self-informative; it is our background knowledge about dogs and the prevalence and duration of explosions that leads us to give it little credence. This is not, after all, a case in which Condition 2 is violated, but rather one in which Condition 2 does not apply because background evidence E contains information relevant to the proposition F in question.

⁴Even granting Condition 2, it may be the case that the Principal Principle admits sufficiently many counterintuitive defeaters as to be unviable. Here is an example. As before, X says that the chance of a positive α -test is 0.8, but now other evidence E , which is not itself a defeater, provides some grounds in favour of a positive f -test: $P(F|XE) = 0.8$, say. One might think that $A \leftrightarrow F$ is a non-defeater,

because it is compatible with both the evidence in favour of F and the chance claim X . But it turns out that $A \leftrightarrow F$ must be a defeater, contrary to any such intuition: F is not a defeater by Proposition 1; if $A \leftrightarrow F$ were also a non-defeater then by Proposition 2, $P(F|XE) = 1/2$, which contradicts the supposition that $P(F|XE) = 0.8$. Other examples of counterintuitive defeat are explored in detail by Wallmann and Hawthorne (2015). We leave it open here as to whether the Principal Principle is generally viable. Our claim is a conditional one: *if* an instance of the Superprincipal Principle is viable, it favours objective Bayesianism over subjectivism.

References

- Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, 24:281–323.
- Lewis, D. K. (1980). A subjectivist's guide to objective chance. In *Philosophical papers*, volume 2, pages 83–132. Oxford University Press (1986), Oxford. With postscripts.
- Meacham, C. J. G. (2010). Two mistakes regarding the Principal Principle. *British Journal for the Philosophy of Science*, 61(2):407–431.
- Wallmann, C. and Hawthorne, J. (2015). Admissibility troubles for direct inference principles. Manuscript, forthcoming.
- Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford University Press, Oxford.