**Undermining Undermined:**
**Why Humean Supervenience Never Needed to be Debugged**
**(Even If It's a Necessary Truth)**[1]

John T. Roberts
University of North Carolina

In the mid-1980s, David Lewis, the great champion of Humean supervenience (henceforth HS), feared that his beloved thesis might fall at the hands of chance. This was because of very strange phenomena that have come to be known as *undermining futures* (Lewis (1986a, xiv-xv)). The existence of undermining futures appears to be implied by any account of chance that is consistent with HS. Lewis's fear was that the existence of undermining futures, together with the Principal Principle -- a thesis he took to be unassailable -- entailed a contradiction. This problem was so troubling to Lewis that he dubbed it "the big bad bug" (Lewis(1986a, xiv)).

There are two strategies that may be employed in attempting to save HS from the bug. The first strategy involves accepting the argument that derives a contradiction from HS together with the Principal Principle, and rejecting the latter. Since, as Lewis argues, the Principal Principle seems to be intimately tied up with our ordinary understanding of chance, such strategies arguably must be somewhat conceptually revisionary. John Halpin (1994), Ned Hall (1994), Michael Thau (1994), and Lewis (1994) himself have advocated versions of this strategy. The other strategy is to try to block the argument used to derive the contradiction. Peter Vranas (1998) has used a version of this strategy. Vranas's objection to the derivation depends crucially on the claim that HS is put forward by Lewis as a contingent truth rather than a necessary one. Vranas's argument successfully rescues HS understood in this way, but it leaves the stronger claim that HS is necessarily true in the clutches of the bug. This stronger version of HS, though, is of interest as well, and many of us actually find it more plausible than Lewis's own view: The considerations that motivate HS in the first place seem to many of us to motivate the claim that HS is necessarily true. In this paper I will argue that the derivation of the contradiction can be blocked without appeal to the contingent status of HS. This strategy will require a slight modification to Lewis's formulation of the Principal Principle, but this modification will not involve any revisionism about our concept of chance. Indeed, I will argue that this modification is motivated by the very reasons Lewis himself gives for accepting the Principal Principle in the first place. The upshot is that even those who view HS as necessarily true, and who do not want to engage in revisionism about the notion of chance, have nothing to fear from the bug.

## 1. Chance and Humean Supervenience

*Chance* is understood here as a probability distribution defined over a set of possible events or a set of propositions. The chance distribution can vary from possible world to possible world, and from time to time within a single possible world. Chance is here taken to be objective. What is meant by the objectivity of chance is a question that I can't treat here. But to a first approximation, the idea is that the chance distribution at a particular world at a particular time is determined by features of that world that are objective in the sense that they do not depend on the

knowledge, beliefs, interests, or preferences of epistemic subjects.

But to say only that is to leave it wide open just what chance is. There are uncountably many different probability distributions that might be defined as functions of the objective features of a possible world at a given time. Only one of these is supposed to be singled out by the label "chance." How is this privileged probability distribution singled out? Lewis has famously proposed a necessary condition on any distribution that deserves the name "chance": The chance distribution is one that is related to credence in a way specified by the Principal Principle:

(PP) Let P be the chance distribution for a given world at a given time. Then at that world at that time, for any reasonable initial credence function C, for any admissible evidence E which specifies the chance of A:

$$C(A/E) = P(A)$$

A "reasonable initial credence function" is a probability distribution that represents the degrees of belief of an ideally rational epistemic subject, whose beliefs would remain reasonable after conditionalizing on new evidence as it came in. Evidence is "admissible" only if it does not contain any "intelligence from the future" concerning how chancy future events will turn out. So the basic idea underlying the PP is simply that the chance distribution is one with the following property: when a cognizer like one of us gets admissible information about this distribution, she ought to set her degrees of belief accordingly. If I know that the chance of my winning the lottery is 1 over 10 billion, and I have no crystal ball which allows me reliably to predict the future, then I ought, rationally, to set my degree of belief that I will win at 1 over 10 billion. To set it any higher or lower would be irrational on my part. According to Lewis, PP is a reasonable principle because it "seems ... to capture all we know about chance" (Lewis (1986b, 86)). I'll return to the issue of why we should accept PP later in this paper.

(Since the supposed problem of undermining futures presupposes the PP, which is phrased in terms of reasonable credence functions, the whole discussion is situated within a broadly Bayesian framework. I will continue working within this framework in this paper. It may well be that Bayesianism is wrong-headed; if so, then much of what follows will be wrong-headed as well, but then so will be the very statement of the problem that I'm trying to solve. If the statement of the problem is actually based on a confusion, then my battle here is already won.)

Humean supervenience about chance is the thesis that the facts about chance are all fixed by a certain range of the world's features. Call this range "the Humean base." More precisely, the thesis might be formulated as follows:

> No two possible worlds agree with respect to their Humean bases but differ with respect to chance.

This is the way some philosophers define "Humean supervenience."[2] If that thesis is true, then it is necessarily true. But Lewis maintains that Humean supervenience is only a contingent truth, true of our world and of all possible worlds in the "inner sphere" of worlds most similar to ours, but not of all possible worlds whatsoever (Lewis (1986a, x)). So Lewis's version of the thesis is not the one above, but rather this one:

> No two possible worlds within the "inner sphere" agree with respect to their Humean bases but differ with respect to chance.

If my argument in this paper is correct, then nothing important to the supposed problem of undermining futures depends on the choice between these two definitions of Humean supervenience.[3]

The "Humean base" referred to is usually defined as the arrangement of local qualities in spacetime. These "local qualities" are supposed to be particular, occurrent features of the world. Such things as mass density, charge density, the presence of hard atoms, the presence of nothing but vacuum, etc. count as local qualities; what is forbidden is such things as global features of the world, dispositions, modal properties, or irreducibly probabilistic properties.

One analysis of objective chance that is consistent with Humean supervenience is the naive frequentist view, according to which the chance that any given event of type T having an outcome O is just the long-run relative frequency with which events of type T have outcome O. A more sophisticated view of chance consistent with Humean supervenience is implied by Lewis's best-system analysis, the details of which I won't go into here (Lewis (1994, 480-2)). On this account, while the chances need not equal the actual long-run frequencies, they cannot vary in a way completely independent of the frequencies. Roughly speaking: frequencies will guide chances, in such a way that other things being equal, a significant difference in frequencies will make for a difference in chances. It is plausible that this kind of connection between frequencies and chances will be implied by any reasonable view of chance consistent with Humean supervenience. For big differences in frequencies are big differences in the Humean base, and differences in the Humean base can make for differences in chances, according to Humean supervenience.

**2. Undermining Futures:**

The phenomenon of undermining futures arises in the following way. The chance distribution at a given world at a given time supervenes not just on its past and present, but on its entire history -- more specifically, on the entire history of its Humean base.[4] It follows that at a typical time, there are multiple possible futures, and some of these possible futures would result in different present chance distributions than the one that actually obtains. An *undermining future* is a possible future that (i) would result in a different present chance distribution than the one that actually obtains, and (ii) currently has a non-zero chance of transpiring.

For example: consider a possible world w, and a typical moment t in the history of w. Suppose that there is a kind of event that occurs a great many times in w -- call it a "coin toss." Coin tosses have two possible outcomes, "heads" and "tails," and suppose that in w, the chance that any particular coin-toss will give the result "heads" is ½. Further suppose that in w there will only ever be finitely many coin-tosses. But the number of coin tosses there will be is astronomical -- say, on the order of Avogadro's number. Finally, suppose that at moment t, only a tiny fraction of all the coin tosses there will ever be have taken place so far.

In world w at moment t, there is a tiny but non-zero chance that *every* future coin toss will turn up "heads." Consider a possible world whose Humean base perfectly matches that of w up to moment t, but in which this surprising long run of heads comes to pass. It seems that on any decent account of chance that is consistent with Humean supervenience, the chance in such a world of getting "heads" on any given coin toss is not ½, but is rather somewhat closer to 1. This is because of the aforementioned way in which chances must tend to follow frequencies if Humean supervenience is correct. So such a long run of heads constitutes an undermining future

for world w at moment t. In w at t, the following is true: Such an astounding future run of heads has a non-zero chance of coming about, but if it were to come about, then the present chance of getting heads on any given toss would be different from what it actually, currently, is. Such a future would undermine the present, actual chance distribution.

The phenomenon of undermining futures is palpably weird. Its weirdness can be brought out in this way: there is currently a non-zero chance that something will happen such that, if it were to happen, then the present chances would have different values from the ones they in fact have. As Lewis writes: "It's not that if this future came about, the truth about the present would change retrospectively. Rather, it would never have been what it actually is, and would always have been something different" (Lewis (1994, 482-3)). So there is currently a non-zero chance that the present chances have values other than the ones they in fact do have. That smells pretty peculiar. And its peculiarity seems to reflect badly on Humean supervenience about chance.

Lewis himself isn't bothered by views with merely peculiar consequences, as his readers are well aware. He is bothered by contradictions, though. And when you put the existence of undermining futures together with PP, a contradiction can be derived. Here's the derivation: Let E be an admissible evidence statement that specifies the present chance of the truth of a proposition F concerning the future; i.e., E entails that the chance of F has some particular value, x. Let F be an undermining future, so that x is not zero, but in any possible world in which F is true, the present chance of F is something other than x. Since E entails that the chance of F is x, and F entails that the chance of F is not x, E and F are inconsistent. So for any reasonable initial credence function C, C(F/E) is zero. But by PP, C(F/E) = x, which is not equal to zero. That's a contradiction.

The apparent upshot is that the existence of chance, Humean supervenience about chance, and PP form an inconsistent triad. This is what has invited the various rescue maneuvers mentioned earlier. I will argue here hat no such rescue maneuver is needed. We need, instead, to recognize that the derivation of the contradiction just rehearsed depends on an inappropriate use of the Principal Principle. The inappropriateness of this use is revealed by a proper understanding of the motivation of this principle.

**3. Dissolving the Problem:**

Let's return to the question of why PP seemed like a good idea in the first place. Rational credences and chances are two totally different kinds of probability; why should we expect there to be any link between them at all? The answer to this question, I think, is the following: even if meaning isn't the same thing as use, our use of chance the notion of chance must place constraints on what can qualify as the referent of our word "chance." Further, it seems that the primary function of the notion of chance in our epistemic practice is this: we let our best estimations of the chances of chancy events guide our degrees of belief about those events. PP is just an attempt to express the way in which we do this.

But it is not an attempt to describe merely our actual applications of the notion of chance. Rather, this attempt involves a certain degree of idealization: chance is related by PP to *reasonable* credence functions, which surely introduce a considerable amount of idealization into the study of our actual epistemic practices. It seems perfectly legitimate to introduce some kind of idealization here. What we refer to by "chance" has to be constrained by our use of the notion

of chance. However, a liberal understanding of our use of the notion of chance can allow that this use includes not just how we actually set our degrees of belief in the light of our best estimates about chances, but also how we take it that we ought, rationally, to set our degrees of belief in the light of these estimates.

The motivation for PP, then, seems to be the plausible idea that we ought to let our best estimate of the chance of a future event guide our degree of belief that that event will occur. But as Lewis himself notes, the formulation of PP given above, which is used in the derivation of the contradiction, expresses only one very special case of the way this guiding takes place.[5] For it applies only to evidence which *specifies* the chance of some event or proposition. Unfortunately, we generally have to make do with evidence that is less informative, evidence which provides some information about the chance of some future event, but that does not entail that this chance has any one particular value.

For example, suppose that I have good evidence that tells me that a certain coin is either fair, or else is rigged so that it will *always* turn up heads. Further, suppose that my evidence makes both of these possibilities seem equally likely. To what degree should I believe that the next toss of this coin will turn up heads? In the circumstances described, I ought to believe to degree ½ that the coin is fair, in which case it has a ½ chance of turning up heads; I ought to believe to degree ½ that the coin is rigged, in which case it has a chance of 1 of turning up heads. So my degree of belief that the coin will turn up heads should be:

½ x ½ + ½ x 1 = 3/4.

The general principle at work here is the following: to figure out the degree to which I ought to believe that a certain chancy future outcome will occur, I should take a weighted average of all the epistemically possible values of the chance of that outcome, where the weight for a particular value is just my rational degree of belief that the chance has that value. Recognizing all this, Lewis writes that the most general form of the Principal Principle is the following:

(GPP) $C(A/E) = \sum_x C(X_x/E)x$

where $X_x$ is the proposition that the chance of A is x (Lewis (1986b, 87)).

So the equation I've labeled "PP" is really just a very special limiting case of the true principle relating chance and credence that, as Lewis writes, "seems .. to capture all we know about chance" (Lewis (1986b, 86)). The true principle, which applies most generally, is what I've labeled "GPP." Furthermore, PP is a special case of GPP that applies only to cases that are impossible in principle. Recall that PP applies to the evidence E only if E *specifies* the objective chance of the proposition A. That means that E does something that real evidence never does: namely, entail that the chance of A has some particular value. In cases in which we reason about chances, we typically possess evidence that gives some degree of confirmation to claims about the chances of various kinds of events occurring under various circumstances, and this kind of evidence can indeed rationally constrain our credences concerning future events. But real evidence *never* constrains these credences by *specifying* the objective chances of such events. Indeed, if Humean supervenience is correct, there could be such evidence only if there were no problem of induction. This is because, on a Humean supervenientist account of chance, the chances supervene on the entire history of the world, including the future. Thus, in order to fall

within the domain of PP, a piece of evidence E would have to *entail* (and not merely probabilify) contingent information about the distant future, something that no evidence in principle available to creatures like us could ever do.

The reason why we ought to believe GPP is this: if rational epistemic practices are to be modeled using probability distributions, then GPP expresses the way in which we ought to make use of our evidence about chance to guide our beliefs about the future. This same reason also motivate the addition of a clause to GPP: GPP applies to evidence E only if E is an evidence proposition that empirical cognizers like ourselves could, in principle, encounter as evidence. This is because GPP is supposed to be a normative principle regulating the rational maintenance of expectations for the future by subjects like us, whose epistemic access to the world is empirical. Such a normative principle should only be expected to apply to situations that such subjects could, in principle, face. It shouldn't be expected to give any answer at all to the question of what such subjects should do in situations that they could not possibly find themselves in.

But if we restrict GPP so that it applies only to cases where the evidence E is evidence that is possible for finite empirical cognizers to obtain, then the *reductio* concerning undermining futures is blocked. This *reductio* works by deriving both that $C(F/E) = 0$ and that $C(F/E) > 0$. But with the restriction just proposed, there is no way to derive that $C(F/E) = 0$. For that equation was derived by appealing to the inconsistency of the evidence E with the proposition F. Without this logical inconsistency, the argument doesn't go through. F is a proposition about the future that is supposed to be logically consistent with the actual occurrent history of the world up to the present. Trivially, no empirical evidence derived from past and present observations could be logically inconsistent with any proposition about the future that is logically consistent with the occurrent history up through the present. So, if E is a piece of evidence that empirical cognizers like ourselves could encounter, or even one that ideally rational empirical cognizers could encounter, then it could not be inconsistent with the proposition F.

The problem here is not simply that the *reductio* makes a simplifying assumption that is dispensable in a more rigorous analysis: without the requirement of inconsistency between the evidence E and the possible future F, the contradiction cannot be derived; it isn't enough for E simply to make F extremely unlikely. Perhaps there is some other way of deriving a contradiction from the existence of undermining futures together with GPP subject to the restriction I've recommended, but there does not appear to be one that works in basically the same way as Lewis's argument. Some totally new argument would be required.

**4. Is There Another problem with GPP-cum-HS?**

How might such a new argument go? In this section I will explore one way of trying to generate a problem for GPP and HS analogous to Lewis's *reductio*, and argue that it won't work.

The GPP says that, for any possible empirical evidence E, and any reasonable initial credence function C, the following relation holds:

(GPP) $C(A/E) = \sum_x C(X_x/E)x$

By probability theory, C must satisfy:

(1) $C(A/E) = \sum_x C(X_x/E)C(A/X_x)$

Hence, C will satisfy GPP iff:

(2) $\sum_x C(X_x/E)x = \sum_x C(X_x/E)C(A/X_x)$.

Suppose that A is an undermining future for a chance distribution that assigns A the non-zero chance z. Then, $C(A/X_z) = 0$, but z is nonzero. So the term in the summation on the right-hand side of (2) corresponding to the value x = z is lower than the corresponding term on the left-hand side. In order for the equality to hold, then, there must be at least some values of x for which the term on the right-hand side is greater than the corresponding term on the left-hand side, that is:

$C(A/X_x) > x$

So GPP, together with the existence of undermining futures, entails that for some propositions A, and for some values of x, $C(A/X_x) > x$, and for other values of x, $C(A/X_x) < x$. Perhaps this consequence can be exploited to derive a contradiction, or perhaps this fact alone, while not contradictory, is implausible enough that the assumptions that lead to it must be rejected.

But it's far from clear that this is so, and in fact there is reason to think that this consequence is plausibly true, given HS about chance. Consider first that the inequalities just derived can only be derived when A is a proposition that could be an undermining future for some present chance distribution. In order for that to be the case, A must be extremely informative. For it must entail that the facts in the Humean base are inconsistent with a certain present chance distribution. (Recall the paradigm case of an undermining future discussed earlier: the proposition that the astronomically many coin-tosses yet to come will all turn up heads.) Since, according to HS, chance supervenes on the *entire* Humean base, any proposition that entailed such a thing would have to carry a tremendous amount of information about the facts in the Humean base. This information would have to be sufficient to rule out many chance distributions that would assign a low chance to A itself.

Now, consider a rational agent, whose epistemic practices are to be modeled by a reasonable credence function. The conditional probabilities of such a function, C(B/D), can be thought of as modeling how much confidence the agent would be willing to grant B, in a hypothetical case in which she accepts D. If the agent were to accept, hypothetically, the proposition $X_x$, where x is a rather high value, this would give her grounds for high confidence in the truth of A deriving from two different sources. The first is simply that $X_x$ assigns a rather high chance to A. Admittedly, it is hard to see why this ground gives her reason to assign confidence to A to a degree any *higher* than x itself. But there is a second ground for confidence in A: $X_x$ declares that the present chance of A is rather high -- specifically, x -- and A is assumed to be informative enough to rule out, definitively, many present chance distributions that would assign lower values to the chance of A. From the agent's point of view, then $X_x$ and A are positively correlated. This is an *additional* ground for confidence in A. Together with the first ground, it provides the agent with a reason to be confident to a degree greater than x itself that A is true, given the hypothesis that $X_x$. This makes it plausible that for a reasonable credence function C, a

proposition A that is informative enough to be an underminer, and a relatively high value of x, it can be the case that $C(A/X_x) > x$. So the fact that some of the terms on the right-hand side of (2) are zero while the corresponding terms on the right-hand side are not does not show that the equality asserted by (2) cannot hold. For these zero terms are those corresponding to values of x that are very low, and the deficit they introduce can be made up for by some terms of the left-hand side corresponding to higher values of x.

On the other hand, consider the case in which A is not informative enough to be an underminer. In this case, there doesn't appear to be any reason why $C(A/X_x)$ could not be equal to x. Indeed, if it is true that:

(4) For any reasonable initial credence function C, and any proposition A that is not informative enough to be an underminer, that $C(A/X_x) = x$,

then (2) will automatically be satisfied. It is tempting to say that (4) states a necessary requirement on initial credence functions. Doing so would vindicate Lewis's original PP for propositions that are not informative enough to be undermining futures, and surely this includes the propositions whose credences we have to evaluate in the course of everyday life and the practice of science. So this requirement would be consistent with GPP, restricted to possible empirical evidence. It would also explain why rational credence functions satisfy the requirement (2), which is entailed by GPP. Furthermore, it would capture our intuitions that in "normal" cases the original PP is just fine -- and it would capture these intuitions in the most straightforward way possible, by implying that PP is *exactly true* in such cases.

**5. The Comparative Virtues of This Solution:**

Lewis's own way of beating the bug involves rejecting the PP in favor of a different principle called NP. For Lewis this involves a price, since PP is, for Lewis, "the key to our concept of chance" (Lewis (1994, 489)). Giving up such a fundamental principle feels like giving up the notion of chance itself! But the price is not too high to bear, in Lewis's estimation, because in "normal" cases, PP is approximately true. NP appears to diverge from PP only in very unusual cases, such as cases of undermining futures, which are not the cases our intuitions about chance are honed on, or the standard cases of usage to which any decent semantics for our term "chance" must be faithful. "[N]ear enough is good enough," writes Lewis (1994, 489). But surely even nearer is even better.

I've argued that the original motivation for the PP motivates only the more general GPP with the qualification that it is only to be applied to evidence statements that could really be empirical evidence. *That* principle, I submit, is the real key to our concept of chance. I've shown that *that* principle, together with HS, are not threatened by the existence of undermining futures, and so we can uphold HS and simultaneously maintain that the key principle concerning our concept of chance is *exactly* true. Furthermore, as the last section showed, this view is consistent with the claim that Lewis's original PP itself is *exactly* true for all cases except those involving propositions informative enough to undermine chance distributions. Even if the price Lewis pays is not excessive, we can get a better bargain; so we should. The availability of a better bargain

has already been demonstrated by Vranas (1988), but his bargain is available only to those who defend HS as a contingent truth. What I've shown here is that defenders of HS as a necessary truth needn't be denied the same benefit.

**Notes:**

1.I would like to thank Joe Camp, John Earman, and Brie Gertler for helpful comments on earlier versions of this paper.

2.E.g., Carroll (1994). Earman's (1984) "Empiricist loyalty test" thesis is essentially Humean supervenience construed in the way just presented in the text.

3.Vranas's (1998) proposed solution to the undermining problem depends crucially on Humean supervenience being construed in Lewis's way. This is what I find troubling about his solution: It seems to me that Humean supervenience construed as a necessary truth is an interesting thesis that should be kept a live option if possible.

4.Lewis explains why this must be so in his (1994, 483-4).

5.Lewis (1986b, 87).